Estimating turning points and testing for climate adaptation *

Tom Bearpark
University of Exeter

Jordan Rosenthal-Kay
Federal Reserve Bank of San Francisco

March 13, 2026

Abstract

Estimating the turning points of nonlinear response functions is central to empirical economics and beyond. We show that the standard turning point estimator suffers from the same pathologies as weak-instrument IV: it is a ratio estimator with a fat-tailed sampling distribution and severe median bias. Standard inference can be misleading, and point estimates are drawn toward the regressor mean. We develop robust sign-restricted confidence sets and shrinkage estimators that pool information across units to stabilise inference. In temperature-GDP and temperature-mortality applications, standard confidence intervals understate uncertainty, and patterns interpreted as climate adaptation are spurious and reflect a statistical artifact.

*Authors' emails: T.Bearpark@exeter.ac.uk, Jordan.Rosenthal-Kay@sf.frb.org. The views in this paper do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System. We thank Gregory Casey, Stephie Fried, Antonio Gasparrini, Tom Hierons, Manuel Linsenmeier, Filippo Palomba, Eric Qian, and Guglielmo Zappalà for helpful comments. We thank Gracie Ellsworth for excellent research assistance. All errors are our own.

1 Introduction

A large empirical literature estimates nonlinear dose-response relationships in which the turning point is a central estimand. Researchers fit quadratic, spline, or binned regression models and use the estimated curve to locate the input level that maximises or minimises the outcome of interest. In climate economics, this turning point determines who 'wins' and 'loses' from global warming (Kotlikoff et al., 2021; Krusell and Smith, 2022; Cruz and Rossi-Hansberg, 2024); in environmental epidemiology, researchers estimate location-specific minimum mortality temperatures and interpret their proximity to local climate normals as evidence of adaptation (Gasparrini et al., 2015; Yin et al., 2019; Tobías et al., 2021; Carleton et al., 2022). Analogous turning-point estimands arise in the study of temperature on GDP growth (Burke et al., 2015; Hsiang, 2025), competition and innovation (Aghion et al., 2005), debt and growth (Reinhart and Rogoff, 2010), agglomeration economies (Au and Henderson, 2006), and many other topics. Turning point estimates also serve as the basis for health guidance on, e.g., the optimal amount of coffee consumption (Crippa et al., 2014) or sleep (You et al., 2024).

In this paper, we show that inference on turning points is statistically fragile for two reasons. First, when curvature is weakly identified, turning point estimators have a *fat-tailed sampling distribution*, analogous to an IV estimator under weak instruments, so conventional delta-method or bootstrapped confidence intervals can severely understate uncertainty. Moreover, these estimators are *median biased* toward the sample mean of the regressor. In environmental economics, the focus of this paper, existing climate impact studies may both overstate the precision of estimated optima—thereby understating uncertainty in both the magnitude and geography of climate damages—and mistake statistical artifacts for evidence of adaptation to climate. To remedy these issues, we develop Anderson-Rubin style confidence sets for turning points from polynomial regressions that have correct coverage, and explore alternative turning point estimators that exploit pooling to attenuate median bias. Theoretically, we focus our attention on turning point estimators from quadratic OLS regressions, for which we are able to provide an exact finite sample characterization, though

we extend our analysis to higher-order polynomials and semi-parametric estimators through simulation. We suggest a simple rule-of-thumb for practitioners: when the t-statistic on (local) curvature falls below 5, confidence set adjustments are necessary and relative median bias exceeds 10%.

To illustrate the basic problem, consider outcomes y_{it} in location i and period t generated by $y_{it} = f_i(T_{it}) + u_{it}$, with $\mathbb{E}[u_{it} \mid T_{it}] = 0$, where T_{it} denotes realized temperature and $f_i(\cdot)$ is twice continuously differentiable and strictly concave. A turning point τ_{0i} exists, defined by $f_i'(\tau_{0i}) = 0$ when $f_i''(\tau_{0i}) \neq 0$. Empirically, researchers estimate f_i and recover an implied optimum $\hat{\tau}_i$ from the first-order condition $\hat{f}_i'(\hat{\tau}_i) = 0$. A local expansion around τ_{0i} yields $\hat{\tau}_i - \tau_{0i} \approx -\frac{\hat{f}_i'(\tau_{0i})}{\hat{f}_i''(\tau_{0i})}$, where $\hat{f}_i''(\tau_{0i}) \to_p f_i''(\tau_{0i})$. The sampling behavior of $\hat{\tau}_i$ depends critically on curvature, the denominator of this ratio, $\hat{f}_i''$. Two implications follow.

First, conventional inference understates uncertainty under weak curvature. When local curvature is weakly identified (e.g., when the t-statistic for $\hat{f}_i''$ evaluated at $\hat{\tau}_i$ is small), $\hat{\tau}_i$ inherits a fat-tailed sampling distribution, analogous to instrumental variables estimators under a weak first stage. Confidence intervals based on local normal approximations undercover (i.e., delta method), and the appropriate uncertainty can be unbounded.[1] Under the null that curvature is zero, the sampling distribution of $\hat{\tau}$ is well-approximated by a Cauchy distribution, centered at the mean of the regressor, which has no finite moments. When an interior optimum τ_0 does exist but curvature is weak, the distribution of $\hat{\tau}$ remains non-Gaussian and is well-approximated by a mixture of (i) this Cauchy component and (ii) a regular component that is approximately normal and centered at τ_0. The mixture weights are governed by the strength of curvature identification, conveniently summarized by the t-statistic on the curvature term.

Second, median bias can generate spurious adaptation patterns. Following the climate impacts literature, we say location i is *fully adapted* to its climate if marginal damages vanish at the prevailing climate, $\mathbb{E}[x_i] = \mu_i$, so that $f_i'(\mu_i) = 0$. When curvature is weakly identified, turning-point

[1] The analogy is to IV under a weak first stage, where the Wald estimator is a ratio of approximately normal statistics and conventional asymptotics fail.

estimators exhibit a systematic bias toward the sample mean of the regressor. In this regime, cross-sectional patterns that resemble climate adaptation can arise mechanically, even when the true underlying response function is identical across locations. Figure 1 illustrates the point. Panel (a) plots country-specific estimates of the turning point of the GDP-temperature response function, $\hat{\tau}_i$, against each country's long-run mean temperature, μ_i, using canonical panel GDP data and a quadratic response function (Burke et al., 2015, see details of the estimation in Appendix A). The figure reproduces a striking empirical regularity: hotter countries appear to have higher estimated turning points, and the relationship lies close to the 45-degree line that would obtain under full adaptation. However, as Panel (b) shows, the same pattern arises in a placebo exercise in which we simulate from a data generating process in which there is no impact of temperature on GDP. Thus, the striking cross-sectional pattern of Panel (a) is spurious. These issues are not unique to a quadratic specification: Appendix Figure A1 repeats this exercise using higher-order polynomials, splines, and binned estimators of the response function, finding the same result. The implication is that correlations between estimated turning points and climate are not, by themselves, informative about adaptation.

These implications are amplified when curvature is allowed to be heterogeneous, since the effective sample size for each estimated optimum is reduced. Researchers face a tradeoff in specifying the degree of heterogeneity that they allow for. Pooling across locations (i.e. imposing $\tau_{0i} = \tau_0$ for all i) yields a more precisely estimated common optimum, since curvature is identified from the full sample, but rules out heterogeneity and precludes the possibility that all units are locally adapted to their climate. Estimating τ_{0i} separately for each unit is more general: it nests adaptation ($\tau_{0i} = \mu_i$), a common optimum ($\tau_{0i} = \tau_0$), and intermediate structures such as $\tau_{0i} = g(X_i)$ for exogenous covariate or grouping variable X_i. Thus, accounting for heterogeneity allows the data to discriminate between different models of adaptation, but this generality requires that curvature be well-identified in each unit individually.

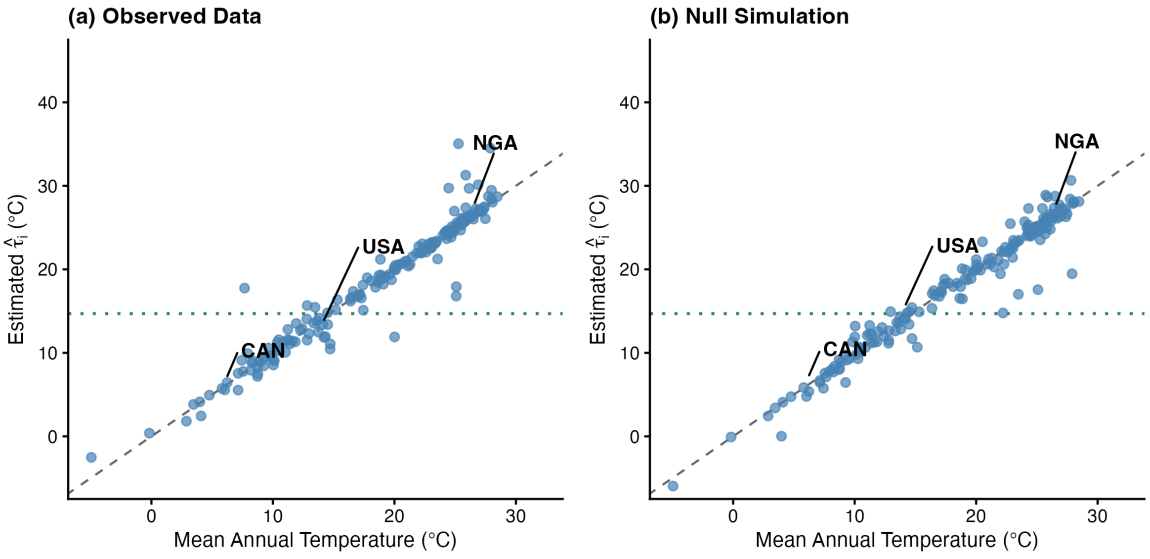


Figure 1: **Spurious adaptation from median bias in turning-point estimates.**
Panel (a) plots country-specific estimates of the turning points, $\hat{\tau}_i$, against each country's long-run mean temperature μ_i using panel GDP data. Details of the estimation procedure used here are in Appendix A.1. Panel (b) repeats the same exercise in which temperature has no effect on GDP, while matching the empirical temperature process and residual variation. Details of the simulation procedure used here are in Appendix A.1. The dashed line is the 45-degree line ($\hat{\tau}_i = \mu_i$), corresponding to full adaptation, and the dotted horizontal line is the pooled optimum estimate.

Contributions

This paper makes three contributions. First, we contribute to the econometrics literature on turning-point statistics and weak IV. We characterize the finite-sample behavior of turning-point (argmax / argmin) estimators when curvature is weakly identified. In a simple parametric setting, we derive the exact sampling distribution of the turning point estimator. We complement our analytic results with simulations showing that the same qualitative pathologies—fat tails, potential unbounded uncertainty, and pronounced median bias toward the regressor mean—arise in semi- and nonparametric implementations commonly used in applied work (Hogan and Schlenker, 2024). While related ratio-estimator issues have been noted in specific applications (e.g., environmental Kuznets curve peaks and elasticity calculations, Hinkley, 1969; Plassmann and Khanna, 2007; Hirschberg et al., 2008; Lind and Mehlum, 2010; Bernard et al., 2015), our contribution is to provide a unified characterization that (i) isolates the *median-bias channel* as a source of spurious cross-sectional

'adaptation' patterns and (ii) connects turning-point estimation directly to weak-identification logic for ratio estimators used in instrumental variables regressions (Staiger and Stock, 1997).

Second, we provide practical solutions for inference and estimation. Building on our diagnosis, we propose two complementary remedies, one for the inference problem, and one for estimation under heterogeneity. For inference, we construct identification-robust confidence sets for turning points by inverting tests of the first-order condition in the spirit of Anderson and Rubin (1949), delivering correct coverage regardless of curvature strength.[2] We also develop an economically-motivated likelihood ratio test that assumes the sign of curvature is known; i.e., whether the turning point reflects a maximum or minimum, and show how to extend our test for turning points estimated from polynomials of finite order. Our tests are therefore closely related to the literature on weak IV with a known first-stage sign (Andrews and Armstrong, 2017; Angrist and Kolesár, 2024), but are, to our knowledge, a novel formulation.

For estimation under heterogeneity, we propose methods that explicitly manage the heterogeneity–identification tradeoff: (i) projection approaches that model optima as functions of observables (e.g., $\tau_{0i} = g(X_i)$), and (ii) a hierarchical Bayesian estimator that shrinks location-specific turning points toward a pooled mean with shrinkage intensity disciplined by local identification strength. Intuitively, heterogeneity is preserved where the data strongly identify curvature, and suppressed where local optima are weakly identified; this yields a disciplined compromise between fully pooled and fully heterogeneous specifications. Conceptually, this approach complements recent Bayesian work on heterogeneous weather effects (Goh et al., 2025).

Third, our work speaks directly to the climate literature and the econometrics of 'adaptation' tests. We apply the framework to canonical temperature–GDP (Burke et al., 2015) and temperature–mortality (Carleton et al., 2022) relationships. Empirically, robust confidence sets for optimal temperatures are often substantially wider than conventional delta-method intervals when curva-

[2]This parallels identification-robust inference for weak-IV settings, where Wald-type approximations can severely undercover (Staiger and Stock, 1997).

ture is weak, and shrinkage-based turning-point estimates can materially alter projected damages relative to standard pooled estimators. More broadly, we provide statistical foundations for using estimated turning points to test for adaptation. This speaks to emerging 'optimum–climate' adaptation diagnostics, like estimating country-specific peaks in spline models (Linsenmeier, 2024), and heterogeneous minimum mortality temperatures common in environmental epidemiology (Gasparrini et al., 2015; Yin et al., 2019; Tobías et al., 2021). It is also related to common two-step exercises that regress estimated response-function features on long-run climate to make inferences about adaptation (Heutel et al., 2021; Auffhammer, 2022).

Moreover, our results have value for inference on turning points in other contexts, for example, in education economics (estimating optimal class size, Kedagni et al., 2021), urban economics (estimating optimal city size, Au and Henderson, 2006), macroeconomics (estimating the turning point in debt-to-GDP/growth relationship, Pattillo, Ricci, et al., 2011; Checherita-Westphal and Rother, 2012; Eberhardt and Presbitero, 2015) and bear upon key findings in industrial organization (Aghion et al., 2005, find that the inverted-U between innovation and competition peaks at the median level of competition).[3] Finally, our results suggest there is substantial under-reported uncertainty in a large epidemiology literature in which 'optimal' amounts of coffee consumption (Crippa et al., 2014), exercise (Schnohr et al., 2015; Arem et al., 2015), screen time (Przybylski and Weinstein, 2017), sleep (You et al., 2024), and so on are reported; claims underlie prominent health guidelines (see, e.g., Vrints et al., 2024).

Outline. The paper proceeds as follows. Section 2 characterizes the sampling distribution of turning-point estimators under weak curvature in a simple parametric setting, and provides simulation evidence that the key pathologies, fat tails and median bias, extend to nonparametric estimators and more realistic data generating processes. We propose robust confidence sets that deliver correct coverage regardless of identification strength, and compare them to standard delta-method

[3]The location of the competition-innovation peak, and how it changes with respect to the technological gap between competitors can be used as a test of Aghion et al. (2005)'s theory; see Hashmi (2013). We return to this example in Appendix C.2.

intervals in temperature–GDP and temperature–mortality applications. Section 3 addresses the heterogeneity–identification tradeoff, developing projection and hierarchical Bayesian approaches that pool information across units while allowing for heterogeneous optima where the data support them. Section 4 concludes.

2 Inference on turning points

In this section, we characterise the econometrics of turning-point estimators. Section 2.1 derives the sampling distribution in a simple parametric setting, showing that it is governed by a curvature-strength parameter θ and ranges from approximately Cauchy (weak curvature) to approximately normal (strong curvature). Section 2.2 provides simulation evidence that these pathologies extend to heteroskedastic settings and to nonparametric estimators. Section 2.3 proposes identification-robust confidence sets that deliver correct coverage regardless of θ, and illustrates the practical difference from delta-method intervals in two canonical applications.

2.1 The sampling distribution of turning point estimators

We begin with the simplest parametric environment in which the distribution of the turning point estimator can be characterized analytically: i.i.d. Gaussian x_i and a quadratic regression with homoskedastic errors, where $\hat{\tau} = -\hat{\beta}_1/(2\hat{\beta}_2)$.[4]

Proposition 2.1 (Sampling distribution of turning points in quadratic models)**.** *Let* $\{(x_i, u_i)\}_{i=1}^n$ *be i.i.d. with*

$$x_i \sim \mathcal{N}(\mu, \sigma_x^2), \qquad u_i \sim \mathcal{N}(0, \sigma_u^2), \qquad x_i \perp u_i,$$

and suppose $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ *with* $\beta_2 \neq 0$*. Define* $\hat{\tau} = -\hat{\beta}_1/(2\hat{\beta}_2)$ *and* $\tau_0 = -\beta_1/(2\beta_2)$.

[4]Note that the presence of fixed effects or additional covariates does not affect the arguments here. We can think of x_i, x_i^2, and y_i as being the residual variation after partialling out such controls. Partialling out fixed effects in models with a non-linear response function does not eliminate the mean of x_i from the identifying variation and thus the results on median bias continue to hold. This is because in models with quadratic or higher-order terms, the mean appears in the residualized higher-order terms (McIntosh and Schlenker, 2006).

1. ***Sampling distribution (Hinkley)*** *For sample size n:*

$$\hat{\tau} \sim \mu - \frac{\sigma_x}{\sqrt{2}} \cdot \frac{Z_1 + \sqrt{2}\left(\frac{\mu-\tau_0}{\sigma_x}\right)\theta}{Z_2 + \theta}, \quad Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1),$$

where

$$\theta = 2\beta_2 \frac{\sigma_x}{\sqrt{2}} \sqrt{n} \frac{\sigma_x}{\sigma_u}.$$

2. ***Behavior indexed by*** $|\theta|$

   (a) Well-identified *($|\theta|$ large): $\hat{\tau}$ is normal, centered at τ_0. Delta-method inference is valid.*

   (b) Poorly-identified *($|\theta|$ small): Heavy-tailed distribution. When both $|\theta|$ and $\frac{\mu-\tau_0}{\sigma_x}$ are small, $\hat{\tau}$ approaches Cauchy$(\mu, \sigma_x/\sqrt{2})$, symmetric about μ (not τ_0). Delta-method inference fails, and the estimator is median biased towards μ.*

3. ***Observable diagnostic*** *For large n, the t-statistic on β_2, $t_{\beta_2} \sim \mathcal{N}(\theta, 1)$, so t_{β_2} estimates θ. (Under normal errors, t_{β_2} is exactly noncentral-t with $n-3$ d.f.; the normal approximation holds for large n.)*

Proof: See Appendix D.1.

Proposition 2.1 shows that the sampling behavior of the turning-point estimator is governed by a parameter, θ, which measures the strength of the curvature signal in the data. The distribution of $\hat{\tau}$ is the Fieller–Hinkley distribution for the ratio of two normals, which can be viewed as ranging continuously from approximately Cauchy when $|\theta|$ is small to approximately normal when $|\theta|$ is large.[5] In Appendix D.2, we provide a complementary asymptotic formulation using local-to-zero sequences for β_2, which delivers the same Hinkley distribution as an exact limit under weaker assumptions (finite fourth moments rather than normality of errors). Appendix D.3 fully derives this distribution in closed form for all values of $|\theta|$. In Appendix D.5, we show that the ratio-of-

[5] Hinkley (1969) derived the distribution for the ratio of two normals on which our proposition is based. Prior to Hinkley, Geary (1930) found a transformation of the ratio of two correlated normals such that the resulting random variable itself was normal, and along the way derives that the ratio of two mean-zero normals is Cauchy distributed, though writes that "this particular case has probably no practical value." Marsaglia (1965) derived the distribution of two normals with unit variance and nonzero means, but incorrectly characterized the ratio of two arbitrary normals. The 1969 paper by Hinkley corrects Marsaglia and notes that his work draws on Fieller (1932), who bounded the error of a normal approximation for the ratio of two normals. The distribution derived in Hinkley (1969) for the ratio of two arbitrary normals, $f(x)$, is heavy tailed in the sense that as $x \to \infty$, $f(x) \approx cx^{-2}$, which implies that it has no finite moments.

normals structure and curvature diagnostic extend to higher-order polynomials and non-Gaussian regressors, but the distribution of $\hat{\tau}$ acquires additional distortion terms that make inference harder.[6]

The appearance of μ in the weak-curvature limit may seem surprising: why should the mean of x matter for a ratio of regression coefficients? Intuitively, with little information about curvature, many quadratic fits achieve nearly identical mean-squared error. OLS resolves this near-indeterminacy by placing the extremum where the data are densest, which is the center of the x distribution. The distribution also depends on $(\tau_0 - \mu)/\sigma_x$, which governs skewness and the speed at which the normal approximation becomes adequate.

Figure 2 illustrates both features. When curvature is weak (left column), the Hinkley distribution is skewed and heavy-tailed, with the direction and magnitude of the skewness governed by $(\tau_0-\mu)/\sigma_x$. When the peak is far from climate (bottom left), the distribution is bimodal and the delta-method normal is visibly displaced from the bulk of the density. When curvature is strong (right column), the normal approximation improves, but convergence is slow when $|\tau_0 - \mu|/\sigma_x$ is large: the turning point lies where data are sparse, generating residual skewness even at $\theta = 3$ (bottom right). The implication for adaptation inference is nuanced. On the one hand, the normal approximation works best when the peak is near climate and curvature is strong — precisely the setting consistent with well-identified adaptation. On the other hand, when curvature is weak, the estimator is most severely biased toward μ when the peak is genuinely near climate, making it impossible to distinguish true adaptation from statistical artifact without accounting for the non-normality of the sampling distribution.

In Appendix Figure A2, we display median bias and confidence interval width as a function of θ for the values of $(\tau_0 - \mu)/\sigma_x$ used in Figure 2. The Hinkley median is a smooth, sigmoid-like function in $|\theta|$, and the Hinkley distribution is median-biased towards μ until θ exceeds 2. The delta

[6]Specifically, two terms that are exactly zero in the quadratic-Gaussian case become nonzero: a correlation correction C_K from the dependence between estimated slope and curvature, and a Taylor remainder driven by f'''/f''. Pooling across heterogeneous climate regimes compounds the problem by inducing a bimodal (or multimodal) regressor distribution, which breaks the slope–curvature independence even for $K = 2$ and shifts the median-bias target from the local regime mean to the global mean.

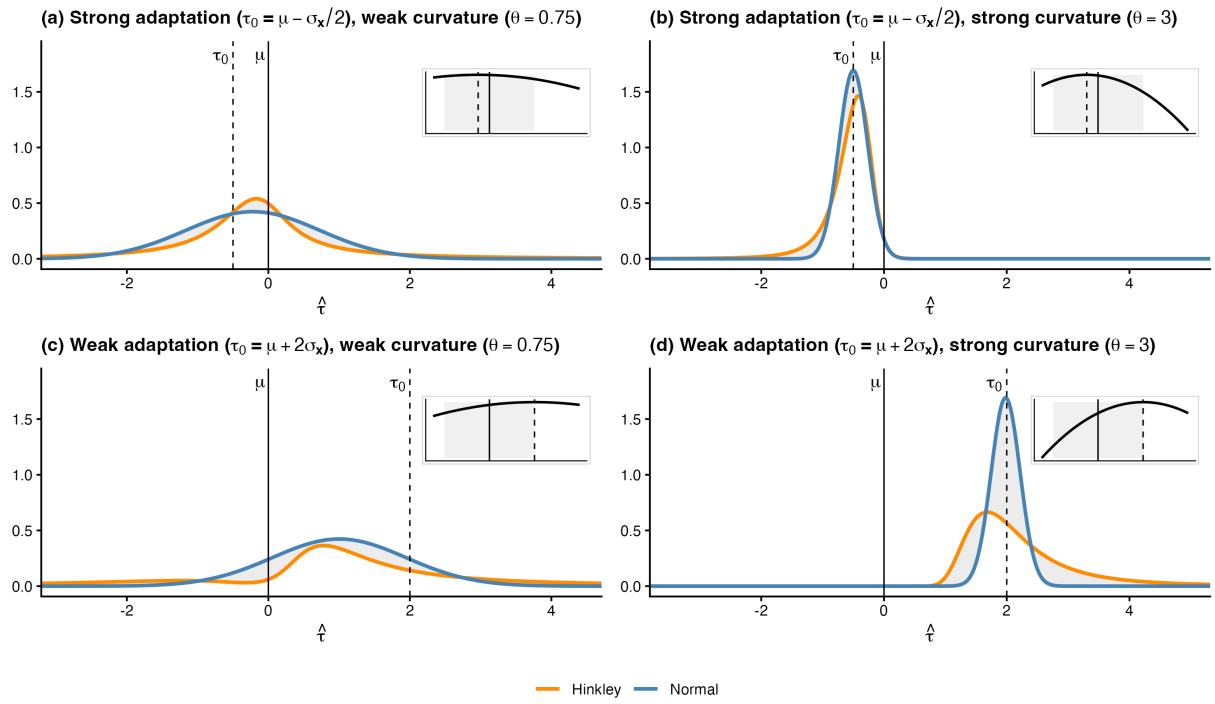


Figure 2: **Sampling distribution of the turning-point estimator.**
Each panel plots the exact Hinkley distribution (blue) against a normal distribution (grey) centered at the Hinkley median with delta-method variance; any discrepancy reflects the failure of the normal shape approximation. Pink shading highlights where the two distributions diverge. Columns vary curvature strength: $\theta = 0.75$ (left, weak) and $\theta = 3$ (right, strong). Rows vary the distance between the true optimum and local climate: $\tau_0 = \mu - \sigma_x/2$ (top, peak near climate) and $\tau_0 = \mu + 2\sigma_x$ (bottom, peak far from climate). Vertical lines mark μ (solid) and τ_0 (dashed). Insets show the underlying response function; the shaded region indicates $\mu \pm 2\sigma_x$: the approximate support of the data. Parameters: $\mu = 0$, $\sigma_x = 1$.

method undercovers for $\theta > 0.5$ and has incorrect coverage when θ is small. When θ is small, the Hinkley distribution is bimodal, so the high-density region is not convex. The confidence interval width shown in the figure therefore does not correspond to a valid confidence set: a single interval necessarily excludes some regions of the support. For this reason bootstrapped confidence intervals will also provide incorrect coverage when curvature is weak.

The extent to which the delta method fails, and the role $(\tau_0 - \mu)/\sigma_x$ plays, can be further understood through power calculations. In the left panel of Appendix Figure A3, we plot power curves for the test $H_0 : \tau_0 = \mu$, as a function of $(\tau_0 - \mu)/\sigma_x$. The delta method performs poorly: when τ_0 is close to μ, even when $|\theta|$ is large, the test is undersized, and rarely rejects the null. As $(\tau_0 - \mu)/\sigma_x$

becomes large, the test becomes oversized, and rejects the null far too frequently, even when $|\theta|$ is large.

2.2 Robustness and extensions

Proposition 2.1 relies on i.i.d. Gaussian regressors, homoskedastic errors, and a correctly specified quadratic model. These restrictions yield closed-form expressions, but are implausible in real world applications. In this subsection, we provide simulation evidence that the core pathologies (median bias toward μ, fat tails, and the failure of normal approximations under weak curvature) survive when these assumptions are relaxed. We then show that they extend to nonparametric estimators, where finite-sample performance is even worse.

Relaxing the DGP assumptions

We simulate $\hat{\tau}$ estimates from the DGP of Proposition 2.1, and then sequentially relax two assumptions: first, we introduce heteroskedasticity; second, we replace the normal distribution for x_i with a skewed distribution, holding the mean and variance fixed at μ and σ_x^2. Appendix Figure A4 reports the results. Introducing heteroskedasticity (middle panel) worsens median bias and generates bimodality and regions of no support when τ_0 differs from μ. Even when the curvature signal is strong ($|t_{\beta_2}| \approx 4$), the sampling distribution remains skewed, with the direction depending on the sign of $\tau_0 - \mu$. Introducing skewness in x_i (bottom panel) has little additional effect: the performance of $\hat{\tau}$ is essentially unchanged. Overall, the qualitative patterns from Proposition 2.1 are robust to these departures from the baseline DGP.

Extension to semi-parametric estimators

The intuition behind Proposition 2.1 extends beyond the quadratic model. Consider a general estimator $\hat{\tau} = \text{argmax}_{x \in \Omega} \hat{f}(x)$, where $\hat{f}$ minimizes a least-squares criterion over some function class $\mathcal{F}$. Provided f is twice differentiable, a Taylor expansion around τ_0 gives $\hat{\tau} - \tau_0 \approx -\frac{\hat{f}'(\tau_0)}{\hat{f}''(\tau_0)}$, so $\hat{\tau}$ is again a ratio estimator, and weak curvature ($\hat{f}''(\tau_0) \approx 0$) generates fat tails even when $\hat{f}'(\tau_0) \to 0$. For parametric estimators with well-behaved $\mathcal{F}$, the argmax continuous mapping theorem delivers $n^{-1/2}$ consistency. However, for nonparametric estimators, where $\hat{f} \to f$ in a function-space norm, the argmax can converge at rates as slow as $n^{-1/3}$ (J. Kim and Pollard, 1990), suggesting substantially worse finite-sample performance.

We examine the behavior of commonly used semi-parametric estimators under a quadratic DGP, relative to our benchmark (OLS with a quadratic specification): a cubic B-spline estimator (6 basis functions), and a binned estimator (10 bins). For each, we simulate the distribution of $\hat{\tau}$ across a grid of signal-to-noise ratios, holding σ_x fixed and varying σ_u. Figure 3 reports the results for two scenarios: a 'strong adaptation' case where τ_0 is near μ (left panel, $\tau_0 = \mu - \sigma_x/2$) and a 'weak adaptation' case where τ_0 is far from μ (right panel, $\tau_0 = \mu + 2\sigma_x$). All three estimators converge to μ as the signal-to-noise ratio deteriorates, confirming that the bias toward μ is not an artifact of the quadratic specification. The quadratic estimator degrades smoothly, transitioning from τ_0 to μ as curvature weakens. The B-spline and binned estimators are more erratic: when τ_0 is far from μ, their finite support acts as a form of implicit regularization, bounding $\hat{\tau}$ within the range of the data. But when τ_0 is near μ and within the data support, the nonparametric estimators exhibit substantially higher variance than OLS. In neither case do flexible methods resolve the weak-identification problem. They merely trade the smooth parametric bias for noisier, less predictable behavior.

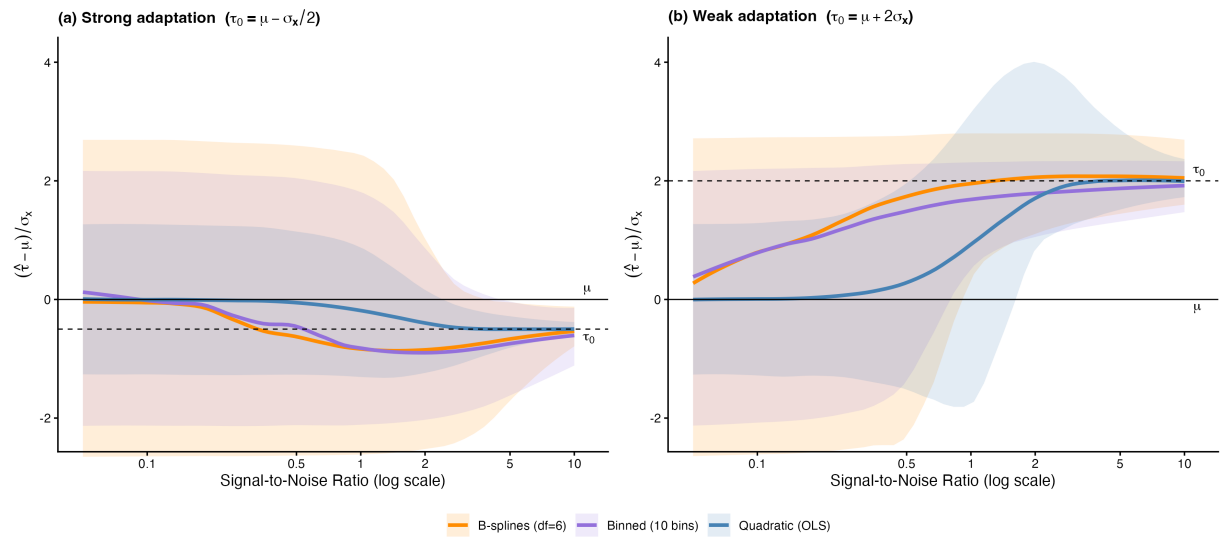


Figure 3: **Neither B-splines nor binning fix weak identification.**
Each panel plots the median estimate of $\hat{\tau}$ (lines) and 68% Monte Carlo confidence bands (shaded areas) across signal-to-noise ratios, defined by σ_x/σ_u (x-axis; holding n fixed), for three estimators: quadratic OLS, cubic B-spline (6 d.f.), and binned (10 bins). Left: $\tau_0 = \mu - \sigma_x/2$ (peak near climate). Right: $\tau_0 = \mu + 2\sigma_x$ (peak far from climate). Horizontal dashed lines mark τ_0 and μ. As the signal weakens (moving right), all three estimators converge toward μ, confirming that median bias is not an artifact of the quadratic specification. Nonparametric estimators exhibit higher variance but no less bias. $n = 200$, 15,000 replications.

Misspecified DGP

Appendix Figure A5 repeats the exercise under misspecified DGPs — an asymmetric quadratic and an asymmetric Gaussian peak — where the true response function is not quadratic. The semiparametric estimators exhibit less bias but substantially higher variance than the quadratic estimator when the model is misspecified, and all estimators continue to show bias toward μ under weak curvature, consistent with the patterns in Proposition 2.1.

2.3 Robust confidence sets for turning points

The results in Section 2.1 imply that delta-method confidence intervals for τ_0 can severely undercover when curvature is weak. We now develop identification-robust alternatives. We first construct

Anderson-Rubin inspired confidence sets by inverting a test of the linear restriction $\beta_1 + 2\beta_2\tau_0 = 0$, which provides correct coverage regardless of θ, extending the weak-IV toolkit of Staiger and Stock (1997) to our setting. We then develop a sign-restricted likelihood-ratio (SLR) test that imposes the economically motivated restriction that τ_0 is a maximum (or minimum), i.e. we treat the sign of β_2 as known. Appendix C extends both constructions to a sign-restricted projection method (SAR) that is more conservative, but has an intuitive geometric interpretation.

Anderson-Rubin inspired confidence sets (AR)

The key observation is that, for any hypothesized value τ, the first-order condition $f'(\tau) = 0$ is equivalent to the linear restriction $H_0(\tau) : \beta_1 + 2\beta_2\tau = 0$. Testing this restriction requires only a standard test on a linear combination of coefficients, which is valid regardless of whether curvature is well-identified. Let $\hat{\Sigma}$ denote a consistent estimator of the variance-covariance matrix of $(\hat{\beta}_1, \hat{\beta}_2)'$. An identification-robust confidence set for the turning point is then obtained by collecting all values of τ that are not rejected:

$$\text{CS}_\alpha^{AR} = \left\{ \tau : \frac{(\hat{\beta}_1 + 2\hat{\beta}_2\tau)^2}{r(\tau)'\hat{\Sigma}r(\tau)} \leq \chi^2_{1,1-\alpha} \right\},$$

where $r(\tau) = (1, 2\tau)$ and $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$-quantile of a chi-squared distribution with one degree of freedom. This construction inverts the test of a linear hypothesis for each candidate τ_0, in the spirit of Anderson and Rubin (1949).

Under the conditions of Proposition 2.1, the set CS_α^{AR} defined above has exact coverage $1-\alpha$ for τ_0, regardless of the value of θ. We confirm this using power calculations in the right panel of Appendix Figure A3, which shows that the Anderson-Rubin test is correctly sized regardless of θ and $(\tau_0 - \mu)/\sigma_x$. When $|\theta|$ is large, CS_α reduces approximately to the standard delta-method interval. When $|\theta|$ is small, the set can be a strict subset of the real line (wide but bounded), the entire real line (reflecting the impossibility of locating τ_0), or the complement of a bounded interval

(an 'exclusive' set indicating where τ_0 is unlikely to be).[7] Note that this test can be constructed for any consistent estimate of the variance-covariance matrix, which includes both cluster- and/or heteroskedasticity-robust estimators.

Sign-restricted likelihood-ratio (SLR) confidence sets

Anderson-Rubin confidence sets provide correct coverage for a turning point, but do not impose whether τ_0 reflects a maximum or minimum of the response function. When curvature is weak, this can be problematic: the confidence set may include values of τ corresponding to $\beta_2 > 0$, which are economically irrelevant if the researcher knows the response is concave. Imposing $\beta_2 < 0$ should exclude these regions and sharpen inference. We now develop a version of the Anderson-Rubin inspired test that imposes this restriction under both the null and alternative.[8] The SLR test asks whether the best-fitting concave quadratic passing through τ fits significantly worse than the best-fitting concave quadratic overall.

Conceptually, the SLR is simply the Anderson-Rubin test with the sign restriction $\beta_2 \leq 0$ imposed symmetrically under both the null and the alternative. Rather than testing $\ell(\tau_0) = 0$ in the unrestricted parameter space, we test it within the cone $\Omega = \{\beta : \beta_2 \leq 0\}$. The test statistic $\omega(\tau)$ measures how much harder it is to fit $\hat{\beta}$ when we impose $\tau = \tau_0$ *relative to* the best fit already satisfying $\beta_2 \leq 0$.

In particular, define the restricted parameter space, $\Omega = \{\beta : \beta_2 \leq 0\}$ and define the parameter

[7] The confidence set CS_α is closely related to the classical Fieller interval for the ratio of two normal means (Fieller, 1954). Fieller's interval for $\beta_1/(2\beta_2)$ inverts the same quadratic inequality that defines CS_α, yielding an equivalent set when the joint distribution of $(\hat{\beta}_1, \hat{\beta}_2)$ is normal. The connection to Anderson–Rubin underscores that the key requirement is linearity of the null in the original parameters, not normality of $\hat{\tau}$ itself. The three geometric cases (i.e. the bounded interval, the entire real line, or complementary set) correspond to the taxonomy of Fieller intervals, governed by whether $g = (c_\alpha/\hat{\theta})^2$ is less than, equal to, or greater than one. When $g > 1$, the curvature estimate is statistically indistinguishable from zero at the α level, and the confidence set is unbounded: no finite turning point can be ruled out.

[8] In Appendix C.1 we develop an alternative procedure which has an intuitive geometric interpretation: the 'sign restricted Anderson Rubin test' (SAR). However, the likelihood ratio based interval described in the main text has superior power.

space under the null $\Omega_0(\tau) = \{\beta : \beta_1 + 2\tau\beta_2 = 0, \beta_2 \leq 0\}$. The likelihood ratio test statistic for a test of $\tau = \tau_0$ under $\beta_2 \leq 0$ is,

$$\omega(\tau) = \min_{\beta \in \Omega_0(\tau)} (\hat{\beta} - \beta)' \hat{\Sigma}^{-1} (\hat{\beta} - \beta) \; - \; \min_{\beta \in \Omega} (\hat{\beta} - \beta)' \hat{\Sigma}^{-1} (\hat{\beta} - \beta)$$

and so,

$$\text{CS}_\alpha^{SLR} = \left\{ \tau : \omega(\tau) \leq \chi^2_{1,1-\alpha} \right\}.$$

When $\hat{\beta}_2 < 0$, the inequality constraints do not bind and $\omega(\tau)$ reduces to the AR statistic, which is χ^2_1 under the null. When $\hat{\beta}_2 \geq 0$, the unrestricted minimum in Ω is attained at the boundary $\beta_2 = 0$, and $\omega(\tau)$ is bounded above by the AR statistic, so the χ^2_1 critical value remains conservative. Formally, the null distribution of $\omega(\tau)$ is $\bar{\chi}^2_{0:1}$, a mixture of χ^2_0 and χ^2_1 (Self and Liang, 1987), so the χ^2_1 critical value provides valid (weakly conservative) coverage.[9]

We can extend this logic to higher-order polynomials. Consider regressions of the form, $y_i = \sum_{k=1}^K \beta_k (x_i)^k + u_i$. Defining $f(x; \beta) = \sum_{k=1}^K \beta_k (x)^k$, we can define the sign-restricted parameter space as $\Omega(\tau) = \{\beta \; : \; f''(\tau; \beta) \leq 0\}$ (where f'' denotes the second x-derivative) and the parameter space under the null as $\Omega_0(\tau) = \{\beta \; : \; f'(\tau, \beta) = 0, f''(\tau, \beta) \leq 0\}$ and compute the likelihood ratio statistic $\omega(\tau)$ as before. Then the α-confidence set for τ_0 under $f''(\tau_0) \leq 0$ simply inverts the test and finds τ values corresponding to $\omega(\tau) \leq \chi^2_{1,1-\alpha}$. For higher-order polynomials, we can define the relevant t-statistic that governs the size of the confidence sets as the t-statistic on, $f''(\hat{\tau}_0, \hat{\beta})$, which is linear in the parameters $\hat{\beta}$.

[9] A tempting alternative is to compute the standard AR confidence set CS_α^{AR} and discard values of τ for which the restricted estimate of β_2 is positive. This "truncated AR" set is a strict subset of CS_α^{AR}, and therefore has coverage weakly below the nominal level. The problem is that the χ^2_1 critical value was calibrated for the unrestricted acceptance region; removing part of that region without adjusting the critical value inflates the rejection probability. The SLR avoids this by placing the sign restriction in both hypotheses: the distance to the restricted alternative (the $\min_{\beta \in \Omega}$ term) acts as a correction that "gives back" what truncation takes away, ensuring that the test has correct size.

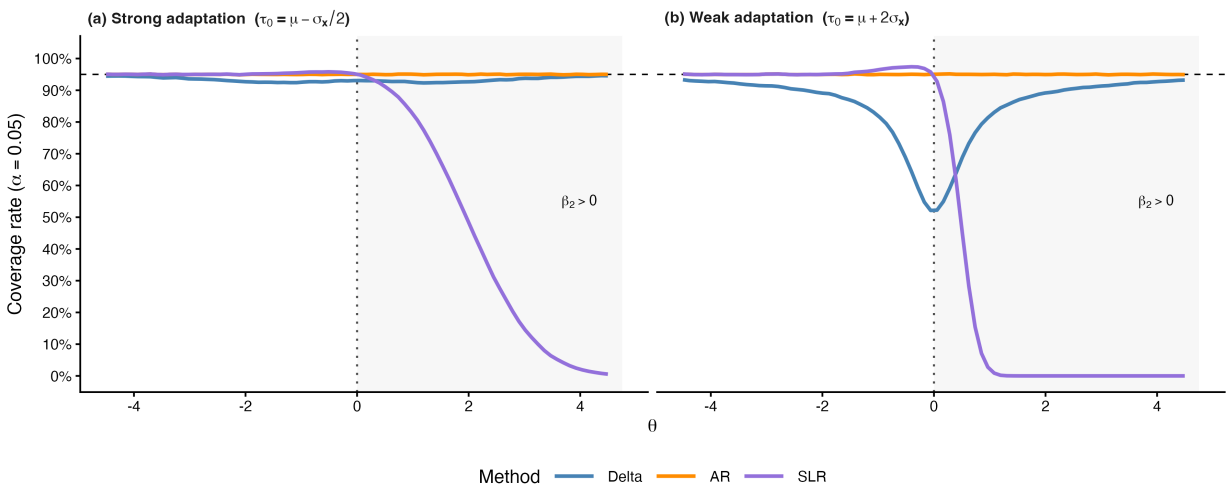


Figure 4: **Coverage of 95% confidence sets for the turning point.**
Each panel plots empirical coverage of nominal 95% confidence sets for the turning point τ_0 as a function of curvature strength θ, under the maintained restriction $\beta_2 < 0$. The horizontal dashed line marks the nominal 95% level; departures below indicate undercoverage. The shaded region ($\theta > 0$) corresponds to $\beta_2 > 0$, where the sign restriction is violated and sign-restricted tests should reject. Left: $\tau_0 = \mu - \sigma_x/2$ (peak near climate). Right: $\tau_0 = \mu + 2\sigma_x$ (peak far from climate). The delta method (blue) undercovers when θ is small and the peak is far from climate. The Anderson–Rubin set (AR, orange) achieves correct coverage throughout. The sign-restricted likelihood ratio set (SLR, purple) matches AR when $\beta_2 < 0$ and rejects when the sign restriction is violated. Based on 20,000 Monte Carlo draws from the DGP of Proposition 2.1.

Comparison of confidence sets and rules-of-thumb

Figure 4 plots empirical coverage of nominal 95% confidence sets as a function of curvature strength θ, under the maintained restriction $\beta_2 < 0$. The delta method undercovers when curvature is weak, and severely so when the peak is far from climate (right panel). The AR set achieves correct coverage throughout, regardless of θ. The SLR set matches AR when the sign restriction is correctly imposed ($\theta < 0$), with slight conservatism near $\theta = 0$ (when a critical value from $\bar{\chi}^2_{0:1}$ is required), and coverage goes to zero as θ grows positive — correctly rejecting turning points that are not maxima.

We now examine if there are reasonable rules-of-thumb researchers can use to ballpark the extent of median bias and delta-method undercoverage in empirical applications. Theoretically, there is no simple rule that determines when these features confound identification and inference, as even

in a simple model they depend on unidentified parameters ($(\tau_0 - \mu)/\sigma_x$ itself is determinative of the size of median bias and AR width), but we can still hope to find a heuristic akin to the first-stage F-statistic ≥ 10 rule of Staiger and Stock (1997). To do so, we generate bootstrap draws from the same DGP underlying Figure 3, and evaluate relative median bias and the relative size of AR and SLR confidence sets for quadratic, cubic, and quartic estimators of the peak as a function of the t-statistic on curvature (local at the peak). Appendix Tables A1 and A2 tabulate the results. AR and SLR confidence sets are unbounded for curvature t-statistics below 1.96, equivalent to the case of a single-instrument first stage F-statistic below 1.96^2.

The $F \approx 10$ rule of thumb from the IV literature (Staiger and Stock, 1997) was itself derived under homoskedasticity; subsequent work has shown that stronger instruments are needed once heteroskedasticity is allowed (Lee et al., 2022). Our setting is analogous: using a t-statistic on curvature of around $\sqrt{10}$ eliminates median bias under the baseline Gaussian-homoskedastic case, but robust confidence sets remain 50% larger than delta-method intervals. In general, our simulations suggest that t-statistics on curvature exceeding 5 are sufficient to both eliminate median bias and ensure true confidence sets are less than 10% wider than those implied by the delta method.

Applications

Figure 5 illustrates the practical implications in two canonical climate applications. Using data from Burke et al. (2015), we estimate a temperature–GDP relationship with strong curvature ($|t_{\hat{\beta}_2}| > 4$); whilst using data from Carleton et al. (2022) we estimate a temperature–mortality relationship with weaker curvature ($|t_{\hat{\beta}_2}| \approx 2.3$) (see details in Appendix A). When curvature is strong, the AR and SLR intervals are only modestly wider than the delta-method interval, and all three largely agree. When curvature is weaker, the gap is stark: the AR and SLR intervals are approximately 2.3 times wider than the delta-method interval, spanning much of the temperature support. In Appendix Figure A7, we compute higher-order response functions and report the corresponding SLR confidence sets for their optima: curvature t-statistics are lower under higher order polynomials

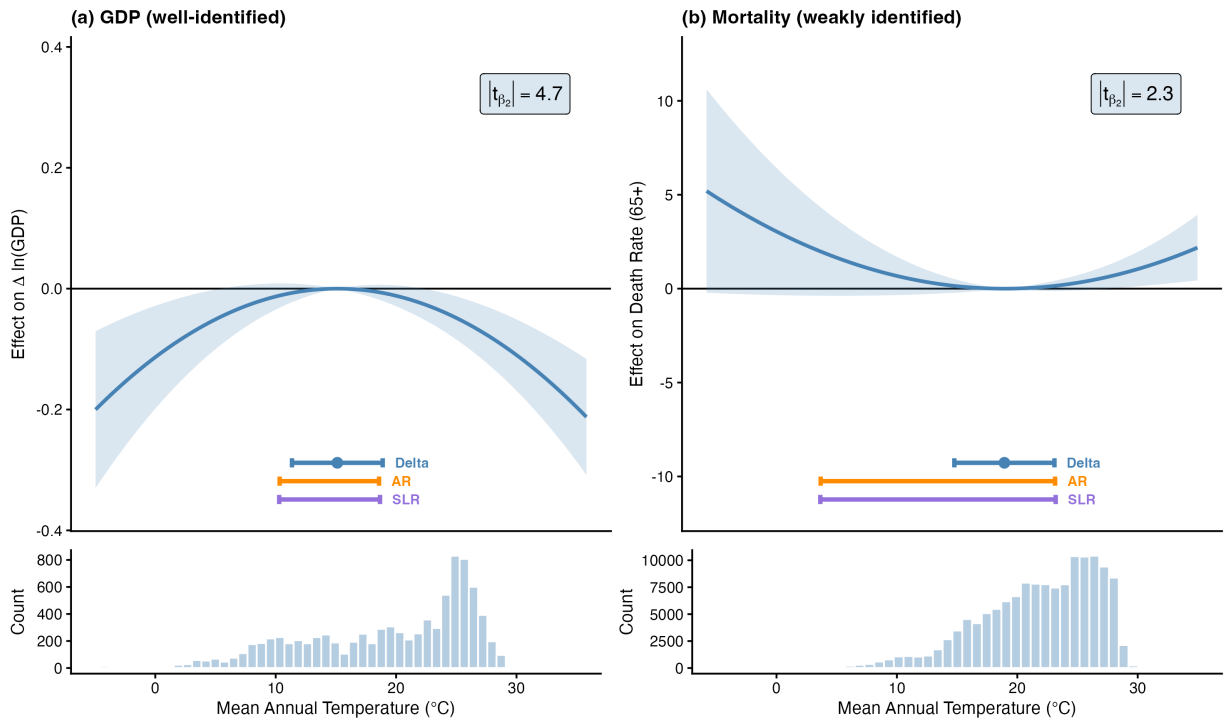


Figure 5: **Delta-method versus weak identification-robust inference on turning points.** *Top row.* Estimated response functions for GDP growth (left, Burke et al., 2015) and 65+ age-group mortality rates (right, Carleton et al., 2022), with pointwise 95% confidence bands. Horizontal bars report three 95% confidence sets for the turning point τ_0: the delta-method interval (blue); the Anderson–Rubin interval (AR, orange), which inverts a χ^2_1 test of $H_0\colon \tau_0 = \tau$; and the sign-restricted likelihood-ratio interval (SLR, purple), which additionally imposes the maintained assumption that the response is concave (GDP) or convex (mortality), recovering a χ^2_1 critical value. We construct our tests using variance-covariance matrix that allows for clustering within countries (GDP) or administrative units (mortality). *Bottom row.* Histograms of temperature data included in the regressions. Further details on data and estimation in Appendix A.

and SLR confidence sets are substantially wider than those implied by the delta method and are disjoint.

Thus, in this example, we cannot reject the hypothesis of full adaptation for most locations in the estimating sample. For the GDP application, where curvature is well-identified ($|t_{\beta_2}| > 4.7$), the delta method and AR confidence intervals are similar in width, and 8 countries fall in the gap between them. For mortality, where curvature is weakly identified ($|t_{\beta_2}| = 2.3$), 31 of the 40 countries in the sample, including most of Northern and Central Europe, have average temperatures that fall outside the delta method confidence interval but inside the AR interval. That is, the delta

method falsely rejects proximity to the estimated optimum for the majority of the sample.

Appendix Figure A6 visualizes how the confidence intervals are constructed, and additionally presents the confidence interval derived from the SAR procedure described in Appendix C.1. For each hypothesized τ_0, we plot the marginal effect $\hat{\beta}_1 + 2\hat{\beta}_2\tau_0$ and its confidence band; the AR set collects all τ_0 for which this band includes zero. When curvature is weak, the marginal effect is nearly flat, so zero cannot be excluded over a wide range and the delta-method interval substantially understates uncertainty about the location of the optimum.

Appendix C.2 shows that the same patterns arise outside climate economics: using data from Aghion et al. (2005), we examine the innovation-maximising level of competition across UK manufacturing industries. Industry-level turning points cluster near mean competitiveness, consistent with median bias, and the AR confidence set is approximately 10% wider than the delta-method interval. This example uses a Poisson regression, rather than the OLS estimator discussed in the main text.

Identification-robust confidence sets tell the researcher which values of τ_0 are consistent with the data, but they do not resolve the median-bias problem identified in Section 2.1: the point estimate $\hat{\tau}$ remains pulled toward μ when curvature is weak. A natural response is to test adaptation directly but such tests have poor power under weak curvature, since detecting departures from adaptation requires precisely estimating β_2.[10] When θ is small, the data cannot distinguish true adaptation from the absence of any curvature. This limitation motivates the estimation approaches developed in Section 3, which address median bias directly by pooling information across units.

[10] E.g., testing $H_0 : \tau_0 = \mu$, which reduces to the linear restriction $\beta_1 + 2\beta_2\mu = 0$. Both AR and SLR confidence sets are likely to cover μ when curvature is weakly identified.

3 Reducing median bias

The AR confidence sets developed in Section 2.3 deliver valid inference regardless of curvature strength, but they do not resolve the point-estimation problem: $\hat{\tau}$ remains median-biased toward μ when θ is small, and this bias cannot be corrected analytically because the Hinkley distribution has no finite moments. The natural alternative is to pool information across units, trading some heterogeneity for improved precision. In this section, we consider two approaches: modeling τ_i as a function of covariates (Section 3.1), which reduces dimensionality but inherits the same weak-curvature pathology; and Bayesian shrinkage (Section 3.2), which regularizes noisy estimates toward a prior. Pooling helps when it is guided by a sensible prior or a genuinely predictive covariate, but data-driven approaches struggle because the Hinkley likelihoods are too diffuse to discipline an empirical prior.

3.1 Modelling τ_i as a function of covariates

One way to reduce the dimensionality of τ_i, and potentially improve its precision, is by modeling it as a function of covariates, rather than allowing it to be fully heterogeneous across units. In this section, we consider linear models of the form, $\tau_i = \bar{\tau} + w_i'\eta$ where w_i is a standardized (mean zero, unit variance) covariate. For example, w_i could capture the level of economic development: e.g., income per capita, as in Carleton et al. (2022), or indicators for development status (Dell et al., 2012; Burke et al., 2015), or could be used to separate weather effects on agricultural and non-agricultural output, as in Conte et al. (2021). This motivates regressions of the form,

$$y_i = \beta_0 + (\beta_1 + \eta w_i)x_i + \beta_2 x_i^2 + u_i. \tag{1}$$

Equation (1) inherits the same weak curvature pitfalls as estimating τ_i with unit-specific heterogeneity: in particular, under the null of no effect, $\tau(w_i)$ estimates will be centered near $\mathbb{E}[x \mid w]$.

Corollary 3.1 explores this property in the same restricted parametric setup as Proposition 2.1.

Corollary 3.1. *Let,*

$$\begin{pmatrix} x_i \\ w_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \rho\sigma_x \\ \rho\sigma_x & 1 \end{pmatrix} \right), \quad i = 1, ..., n.$$

Consider a regression of some random variable y_i, $y_i = \beta_0 + (\beta_1 + \eta_1 w_i)x_i + \beta_2 x_i^2 + u_i$ *and suppose* $\mathbb{E}[u \mid x, w] = 0$ *and* $Var(u \mid x, w) = \sigma_u^2$. *The statistic* $\tau(w) = -\frac{1}{2}\frac{\hat{\beta}_1 + \hat{\eta} w}{\hat{\beta}_2}$ *has the following property for OLS estimators of* $\hat{\beta}, \hat{\eta}$*: When* $\beta_1 = \beta_2 = \eta_1 = 0$,

$$\tau(w) \sim \textit{Cauchy}\left(\breve{\mu}(w), \breve{\sigma}(w)\right)$$

where,

$$\breve{\mu}(w) = \lambda \frac{\mathbb{E}[x \mid w]}{1 + \rho^2} + (1 - \lambda)\,\mu,$$

$$\breve{\sigma}(w) = \frac{\sigma_x}{\sqrt{2}} \times \frac{1}{1 + \rho^2} \times \sqrt{1 + \rho^2(1 - 2\lambda)} \times \sqrt{1 + \rho^2 + \lambda\left(w - \rho\frac{\mu}{\sigma_x}\right)^2}$$

and $\lambda = \frac{1+\rho^2}{(\mu/\sigma_x)^2(1-\rho^2)+(1+\rho^2)}$.

Proof. See Appendix D.4. □

The main idea of Corollary 3.1 is that modeling the peak as a linear function of a covariate w_i can improve precision on β_2, but under weak curvature, the sampling distribution inherits the same unattractive features of the distribution outlined in Proposition 2.1: it is fat-tailed and median biased. However, it is median biased towards a convex combination of a ρ-adjusted $\mathbb{E}[x \mid w]$ and μ, with weights dependent on the correlation ρ.

Often, researchers include interactions between w_i and x_i^2 in the regression as well, so implicitly the peak is, $\tau(w) = -\frac{1}{2}\frac{\beta_1 + \eta_1 w}{\beta_2 + \eta_2 w}$ (where η_2 is the coefficient on $x_i^2 \cdot w_i$). Likewise, this is Cauchy under the null, but its location parameter is not an easily interpretable function of $\mathbb{E}[x \mid w]$. Appendix Figure A9 plots the Cauchy median $\breve{\mu}(w)$ for models in which both the peak τ_i and the slope γ_i are modeled as linear functions of w_i.[11] The plot shows that when including the $w_i x_i^2$ interaction, $\breve{\mu}(w)$

[11] Modeling the slope as a linear function means including $w_i x_i^2$ as a regressor.

behaves like $\mathbb{E}[x \mid w]$ for over most of the support of w_i, but for outlier values of w, the bias tends towards μ. When there is no relationship between w_i and x_i, i.e., $\rho = 0$, $\check{\mu}(w) = \mu$, regardless of the model.

We implement this approach in which w_i represents the standardized mean log GDP per capita at the country level (for temperature-GDP) and at the admin 1 level (for mortality). Appendix Table A3 reports regression results in which w_i is interacted with x_i and its square. Appendix Figure A10 visualizes the implied peaks from these regressions, plotting $\hat{\tau}(w)$ with our SLR confidence bands, as well as the pooled estimate $\hat{\tau}$ and $\check{\mu}(w)$.

In the temperature-GDP relationship, we find that estimated $\hat{\tau}(w)$ are very imprecise and thus statistically indistinguishable from both the pooled estimate and the null median $\check{\mu}(w)$. Including the interaction $x_i^2 w_i$ significantly widens the confidence sets for $\hat{\tau}(w)$, making them unbounded for large w. However, for mortality, including w_i introduces meaningful τ_i heterogeneity that *is* statistically distinguishable from both the pooled estimate and $\check{\mu}(w)$ in poorer countries. Both linear and quadratic estimates of $\hat{\tau}(w)$ slope down: in poorer countries, the minimum mortality temperature is higher. As poorer countries are hotter at baseline, we take this as evidence of adaptation.

3.2 Bayesian approaches

A complementary approach is to treat τ_i as drawn from a distribution and shrink unit-specific estimates toward a common prior. In this subsection, we consider two implementations of this approach. We focus on a Bayesian approach in which we shrink towards the pooled τ baseline and a modern (NPMLE) empirical Bayes approach in which we non-parametrically estimate the prior.

Fixed prior: the pooled sampling distribution

We begin with a simple approach in which we treat the pooled (across all units) estimate of τ as a prior. We estimate the pooled model with country and year fixed effects, draw 20,000 parametric bootstrap samples of $(\hat{\beta}_1, \hat{\beta}_2)$ from the asymptotic joint normal distribution, compute $\hat{\tau} = -\frac{1}{2}\hat{\beta}_1/\hat{\beta}_2$ for each draw, and smooth the resulting sampling distribution using a Gaussian mixture. This smoothed density serves as a common prior for each unit's τ_i. We use this bootstrapped prior, as opposed to the analytic Hinkley, as the data do not match the assumptions of Proposition 2.1 exactly: the errors are not homoskedastic (we use heteroskedasticity-consistent standard errors) and time-series temperature data is non-normal. Regardless, the prior resembles a near-normal Hinkley (here, $|\theta| \approx 4.5$). For each country, we similarly bootstrap from the country-level joint coefficient distribution to approximate the Hinkley-like likelihood, smooth it with a Gaussian mixture, and obtain the posterior by multiplying prior and likelihood on a fine grid of candidate peaks, normalizing numerically. We report posterior means and 95% credible intervals.[12] We additionally shrink $\hat{\gamma}_i$ estimates towards the pooled estimate γ. As $\hat{\gamma}_i$ is asymptotically normal, we use normal-normal shrinkage.

Figure 6 reports the results for the temperature–GDP relationship. Panel (a) plots posterior means against each country's mean temperature. The estimator shrinks weakly identified countries—especially those with extreme temperatures—toward the pooled estimate, while well-identified countries retain peaks closer to their OLS estimates. The 45-degree line (the approximate null under weak curvature, where $\hat{\tau} \approx \bar{x}_i$) and the pooled τ line provide visual benchmarks: posterior means cluster nearer the pooled line than the 45-degree line, indicating effective shrinkage away from the spurious correlation. Appendix Figure A11 visualizes the shrinkage country by country: for each country ordered by mean temperature, it plots the OLS and posterior mean $\hat{\tau}_i$, connected by

[12] We estimate BIC-selected Gaussian mixtures and fit the empirical density well. However, the ability to take means here is not an artifact of Gaussian smoothing. The regularization comes from the Bayesian updating itself: while Hinkley distributions have no finite moments, the product of two Hinkleys has tails that behave like x^{-4}, which allows for the existence of a first (and second) moment.

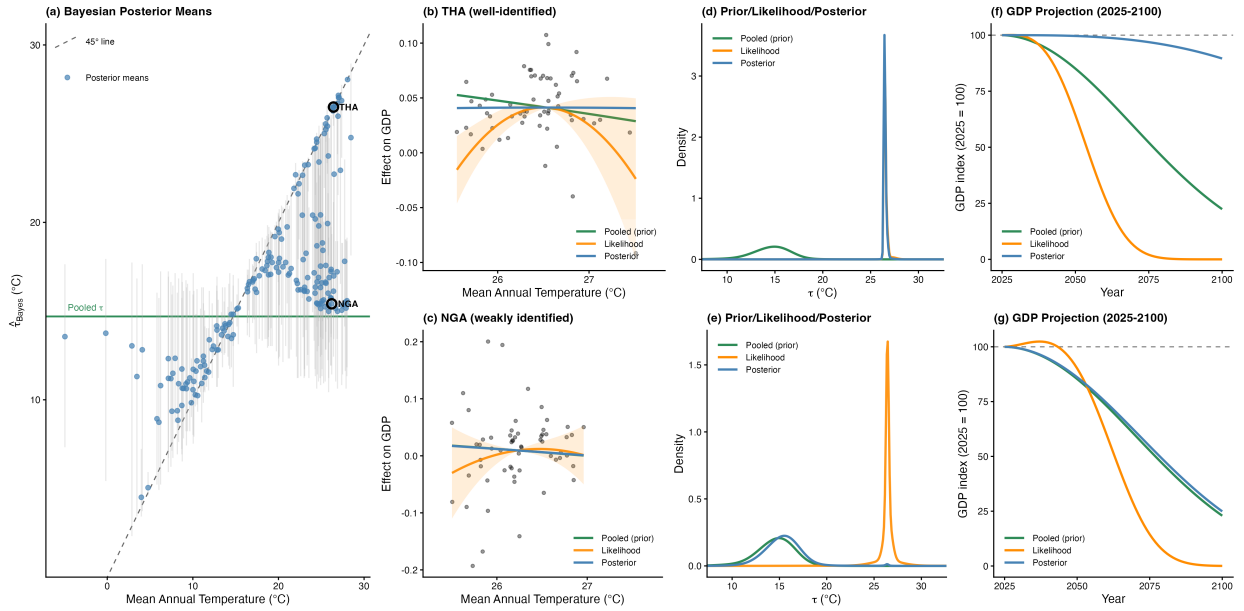


Figure 6: **Bayesian shrinkage toward the pooled estimate in the temperature–GDP relationship.**
Panel (a): posterior mean $\hat{\tau}_{\text{Bayes}}$ against country mean temperature, with 95% posterior credible intervals (grey bars). The dashed line is the 45-degree line ($\hat{\tau} = \bar{x}$, the approximate null under weak curvature); the solid green line is the pooled $\hat{\tau}$. Posterior means cluster near the pooled line rather than the 45-degree line, indicating effective shrinkage away from the spurious correlation. Circled points are the example countries in panels (b)–(g). *Panels (b)–(c):* scatter of GDP growth against temperature with pooled (green) and country-specific (orange) quadratic fits for a well-identified country (Thailand, $|t_{\beta_2}| > 2$) and a weakly identified country (Nigeria, $|t_{\beta_2}| < 1.5$). *Panels (d)–(e):* prior (pooled sampling distribution, green, using a variance-covariance matrix that clusters by panel unit), likelihood (country-specific Hinkley likelihood, orange), and posterior (blue) densities for the same two countries. For Thailand, the likelihood dominates; for Nigeria, the prior dominates. *Panels (f)–(g):* projected GDP index (2025 = 100) under uniform $+3$°C warming from 2025–2100, comparing posterior (blue) and pooled (green) damage estimates.

segments indicating the direction and magnitude of the shift toward the pooled estimate. Panels (b)–(c) illustrate the mechanism for two example countries. For Thailand (well-identified, $|t_{\beta_2}| > 2$), the country-level likelihood is concentrated and dominates the prior, so the posterior closely tracks the likelihood (panel d). For Nigeria (weakly identified, $|t_{\beta_2}| < 1.5$), the likelihood is diffuse and the prior dominates, pulling the posterior toward the pooled estimate (panel e). Panels (f)–(g) translate these estimates into GDP projections under a uniform $+3$°C warming scenario from 2025–2100: for well-identified countries, the posterior and pooled projections diverge meaningfully, reflecting heterogeneity in peaks; for weakly identified countries, the posterior projection collapses toward the pooled estimate. GDP projections based solely on the country-specific OLS estimates imply

implausibly large damages that are considerably attenuated through Bayesian shrinkage.

We repeat this exercise for the temperature–mortality relationship, estimating admin-1-level heterogeneity in the minimum-mortality temperature for the 65+ age group. Appendix Figure A12 reports the results. The same pattern emerges: well-identified admin-1 units retain posterior means close to their OLS estimates, while weakly identified units are pulled toward the pooled minimum-mortality temperature, attenuating the spurious correlation between $\hat{\tau}_i$ and local climate normals.

Data-driven prior: nonparametric empirical Bayes

The fixed-prior approach imposes a specific view about the distribution of τ_i across units. An alternative is to let the data determine the prior using empirical Bayes methods.[13] Standard normal-likelihood, normal-prior models are inappropriate here because the sampling distribution of $\hat{\tau}_i$ is heavily skewed. Instead, we use nonparametric maximum likelihood estimation (NPMLE) of the prior G, modeling it as a discrete mixture over a fine mesh of candidate peaks: $G(\tau) = \sum_{j=1}^{J} \phi_j \delta(\tau - \tau_j)$, where the mixture weights $\phi_j \geq 0$ are recovered by solving a convex program. Appendix E details this approach.

The estimated NPMLE prior is essentially uninformative, placing nearly uniform mass across the observed temperature support, for both GDP and mortality. Because each unit's Hinkley likelihood is so diffuse under weak curvature, the data provide little information about the cross-sectional distribution of τ_i. The NPMLE cannot distinguish genuine heterogeneity in peaks from the noise inherent in weakly identified ratio estimators. Consequently, the posterior means under the NPMLE prior exhibit minimal shrinkage, and the median bias toward μ_i persists.

The contrast between the two Bayesian approaches underscores the identification–estimation trade-off at the heart of this paper. When the researcher is willing to impose structure, such as a fixed prior

[13] There are many interpretations of these estimators: e.g., to estimate a hierarchical model, to solve for decision rules that minimize compound risk, or to mimic an oracle. See Efron (2019) and Koenker and Gu (2024) or Walters (2024) for exposition.

that pools toward a common peak, shrinkage is effective and produces economically meaningful estimates of partial adaptation. When the prior is left to the data, the weak-curvature likelihoods are too diffuse to learn from, and the bias remains. This parallels the finding in Section 3.1: covariates help when they genuinely predict τ_i, but of course cannot overcome the fundamental precision problem created by weak curvature.

4 Conclusion

This paper has shown that turning-point estimators in quadratic regressions inherit the pathologies of ratio statistics when curvature is weakly identified: fat-tailed sampling distributions, median bias toward the regressor mean, and conventional confidence intervals that dramatically understate uncertainty. The strength of these pathologies is governed by a single parameter, θ, which is consistently estimated by t_{β_2}, the t-statistic on the quadratic term. These findings are not artifacts of parametric assumptions: they persist under heteroskedasticity, non-normal regressors, and semi-parametric estimation, for both quadratic and other DGPs.

Our results carry direct implications for applied work. At a minimum, researchers should report t_{β_2} alongside any turning-point estimate as a diagnostic for identification strength, and should use identification-robust confidence sets (which we construct by inverting tests of the first-order condition) rather than delta-method or bootstrapped confidence intervals. When heterogeneous optima are of interest, shrinkage toward a pooled estimate can reduce median bias, but only when the researcher is willing to impose structure on the prior; purely data-driven approaches cannot overcome the fundamental precision problem created by weak curvature. These recommendations apply not only in climate economics but wherever turning-point estimation arises: optimal class size, optimal city size, the debt-growth relationship, or the competition-innovation relationship of Aghion et al. (2005) (see discussion in Appendix C.2).

In the temperature–GDP application, the pooled relationship is well identified ($|t_{\beta_2}| \approx 4.7$), and the pooled optimal temperature is relatively precisely estimated. But once curvature is allowed to vary across countries or subgroups, most units have weakly identified curvature, and the striking cross-sectional pattern in which hotter countries appear to have higher optimal temperatures (Figure 1) is statistically indistinguishable from a placebo exercise in which temperature has no effect on GDP. Interacting the peak with GDP per capita does not resolve this: estimated $\hat{\tau}(w)$ cannot be distinguished from the pooled estimate for most of the income distribution. Bayesian shrinkage toward the pooled peak is more successful, producing estimates consistent with partial adaptation, but the data lack the power to determine the degree of heterogeneity without imposing prior structure. The temperature–mortality application tells a broadly similar story, with one important additional finding: even the pooled minimum-mortality temperature is only moderately identified ($|t_{\beta_2}| \approx 2.3$), so the Anderson–Rubin confidence set for the pooled optimum spans much of the temperature support. This is a cautionary finding for the large environmental epidemiology literature that estimates location-specific minimum-mortality temperatures and interprets their proximity to local climate as evidence of adaptation.

Several directions for future research remain open. The Bayesian shrinkage estimator we outlined in Section 3.2 uses a simple fixed prior derived from the pooled model; richer hierarchical specifications may improve performance, and covariates and Bayesian approaches may be combined. Our empirical applications to temperature–GDP and temperature–mortality are intended to illustrate the statistical issues rather than to deliver definitive damage estimates. More broadly, developing formal decision rules for when heterogeneous turning-point estimation is feasible would provide useful guidance for applied researchers. In the case of quadratic regression, we have shown that this object is the t-statistic on the quadratic term. It would be valuable to define the analogous object for semi- and non-parametric estimators. Such tools would be valuable not only in climate economics but in any empirical setting where turning points are central estimands, since the weak-curvature pathologies we document apply whenever the curvature term is imprecisely estimated. Finally, our results suggest that the climate impacts literature may benefit from greater formalism

about what constitutes evidence for climate adaptation. Throughout this paper, our definition of adaptation more closely resembles acclimatisation (the optimal temperature is the average) than, e.g., resilience (reduced sensitivity to variance). Formalizing the hypotheses being tested, the estimands of interest, and the identification requirements for distinguishing adaptation from statistical artifacts would complement the substantive insights of this literature and strengthen the empirical foundations for climate impacts research.

References

Aghion, Philippe et al. (2005). "Competition and innovation: An inverted-U relationship". In: *The Quarterly Journal of Economics* 120.2, pp. 701–728.

Anderson, T. W. and Herman Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations". In: *The Annals of Mathematical Statistics* 20.1, pp. 46–63. DOI: 10.1214/aoms/1177730090.

Andrews, Isaiah and Timothy B Armstrong (2017). "Unbiased instrumental variables estimation under known first-stage sign". In: *Quantitative Economics* 8.2, pp. 479–503.

Angrist, Joshua and Michal Kolesár (2024). "One instrument to rule them all: The bias and coverage of just-ID IV". In: *Journal of Econometrics* 240.2, p. 105398.

Arem, Hannah et al. (2015). "Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship". In: *JAMA internal medicine* 175.6, pp. 959–967.

Au, Chun-Chung and J Vernon Henderson (2006). "Are Chinese cities too small?" In: *The Review of Economic Studies* 73.3, pp. 549–576.

Auffhammer, Maximilian (2022). "Climate Adaptive Response Estimation: Short and long run impacts of climate change on residential electricity and natural gas consumption". In: *Journal of Environmental Economics and Management* 114, p. 102669.

Bernard, Jean-Thomas et al. (2015). "Environmental Kuznets curve: Tipping points, uncertainty and weak identification". In: *Environmental and Resource Economics* 60.2, pp. 285–315.

Burke, Marshall, Solomon Hsiang, and Edward Miguel (2015). "Global non-linear effect of temperature on economic production". In: *Nature* 527.7577, pp. 235–239.

Carleton, Tamma et al. (2022). "Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits". In: *The Quarterly Journal of Economics* 137.4, pp. 2037–2105.

Checherita-Westphal, Cristina and Philipp Rother (2012). "The impact of high government debt on economic growth and its channels: An empirical investigation for the euro area". In: *European Economic Review* 56.7, pp. 1392–1405.

Conte, Bruno et al. (2021). “Local sectoral specialization in a warming world”. In: *Journal of Economic Geography* 21.4, pp. 493–530.

Crippa, Alessio et al. (2014). “Coffee consumption and mortality from all causes, cardiovascular disease, and cancer: a dose-response meta-analysis”. In: *American journal of epidemiology* 180.8, pp. 763–775.

Cruz, José-Luis and Esteban Rossi-Hansberg (2024). “The economic geography of global warming”. In: *Review of Economic Studies* 91.2, pp. 899–939.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken (2012). “Temperature shocks and economic growth: Evidence from the last half century”. In: *American Economic Journal: Macroeconomics* 4.3, pp. 66–95.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1, pp. 1–22.

Eberhardt, Markus and Andrea F Presbitero (2015). “Public debt and growth: Heterogeneity and non-linearity”. In: *Journal of International Economics* 97.1, pp. 45–58.

Efron, Bradley (2019). “Bayes, oracle Bayes and empirical Bayes”. In: *Statistical Science* 34.2, pp. 177–201.

Fieller, Edgar C (1932). “The distribution of the index in a normal bivariate population”. In: *Biometrika* 24.3/4, pp. 428–440.

— (1954). “Some problems in interval estimation”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 16.2, pp. 175–185.

Gasparrini, Antonio et al. (2015). “Mortality risk attributable to high and low ambient temperature: a multicountry observational study”. In: *The Lancet* 386.9991, pp. 369–375.

Geary, Robert Charles (1930). “The frequency distribution of the quotient of two normal variates”. In: *Journal of the Royal Statistical Society* 93.3, pp. 442–446.

Goh, Gyuhyeong et al. (2025). "Bayesian space–time varying coefficient modeling for climate econometrics: A spatial–temporal Gaussian process approach". In: *Journal of Econometrics*, p. 106075.

Hashmi, Aamir Rafique (2013). "Competition and innovation: The inverted-U relationship revisited". In: *Review of Economics and Statistics* 95.5, pp. 1653–1668.

Heutel, Garth, Nolan H Miller, and David Molitor (2021). "Adaptation and the mortality effects of temperature across US climate regions". In: *Review of Economics and Statistics* 103.4, pp. 740–753.

Hinkley, David V (1969). "On the ratio of two correlated normal random variables". In: *Biometrika* 56.3, pp. 635–639.

Hirschberg, JG, JN Lye, and DJ Slottje (2008). "Inferential methods for elasticity estimates". In: *Journal of Econometrics* 147.2, pp. 299–315.

Hogan, Dylan and Wolfram Schlenker (2024). "Chapter 2 - Empirical approaches to climate change impact quantification". In: ed. by Lint Barrage and Solomon Hsiang. Vol. 1. Handbook of the Economics of Climate Change. North-Holland, pp. 53–111. DOI: 10.1016/bs.hesecc.2024.10.006.

Hsiang, Solomon (2025). *The Global Economic Impact of Climate Change: An Empirical Perspective*. Working Paper 34357. National Bureau of Economic Research. DOI: 10.3386/w34357.

Kedagni, Desire et al. (2021). "Does class size matter? How, and at what cost?" In: *European Economic Review* 133, p. 103664.

Kim, Jeankyung and David Pollard (1990). "Cube root asymptotics". In: *The Annals of Statistics*, pp. 191–219.

Kim, Youngseok et al. (2020). "A fast algorithm for maximum likelihood estimation of mixture proportions using sequential quadratic programming". In: *Journal of Computational and Graphical Statistics* 29.2, pp. 261–273.

Koenker, Roger and Jiaying Gu (2024). "Empirical Bayes for the reluctant frequentist". In: *arXiv preprint arXiv:2404.03422*.

Koenker, Roger and Ivan Mizera (2014). "Convex optimization, shape constraints, compound decisions, and empirical Bayes rules". In: *Journal of the American Statistical Association* 109.506, pp. 674–685.

Kotlikoff, Laurence et al. (2021). "Making Carbon Taxation a Generational Win Win". In: *International Economic Review* 62.1, pp. 3–46. DOI: 10.1111/iere.12483.

Krusell, Per and Anthony A. Smith (2022). *Climate change around the world*. Working Paper.

Lee, David S. et al. (2022). "Valid t-Ratio Inference for IV". In: *American Economic Review* 112.10, pp. 3260–90. DOI: 10.1257/aer.20211063. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20211063.

Lind, Jo Thori and Halvor Mehlum (2010). "With or without U? The appropriate test for a U-shaped relationship". In: *Oxford bulletin of economics and statistics* 72.1, pp. 109–118.

Linsenmeier, Manuel (2024). "Global variation in the preferred temperature for recreational outdoor activity". In: *Journal of Environmental Economics and Management* 127, p. 103032.

Marsaglia, George (1965). "Ratios of normal variables and ratios of sums of uniform variables". In: *Journal of the American Statistical Association* 60.309, pp. 193–204.

McIntosh, Craig T and Wolfram Schlenker (2006). "Identifying non-linearities in fixed effects models". In: *UC-San Diego Working Paper*.

Pattillo, Catherine, Luca Antonio Ricci, et al. (2011). "External debt and growth". In: *Review of Economics and Institutions* 2.3, p. 30.

Plassmann, Florenz and Neha Khanna (2007). "Assessing the precision of turning point estimates in polynomial regression functions". In: *Econometric Reviews* 26.5, pp. 503–528.

Przybylski, Andrew K and Netta Weinstein (2017). "A large-scale test of the goldilocks hypothesis: quantifying the relations between digital-screen use and the mental well-being of adolescents". In: *Psychological science* 28.2, pp. 204–215.

Reinhart, Carmen M. and Kenneth S. Rogoff (2010). "Growth in a Time of Debt". In: *American Economic Review* 100.2, pp. 573–578.

Schnohr, Peter et al. (2015). "Dose of jogging and long-term mortality: the Copenhagen City Heart Study". In: *Journal of the American College of Cardiology* 65.5, pp. 411–419.

Self, Steven G and Kung-Yee Liang (1987). "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions". In: *Journal of the American Statistical Association* 82.398, pp. 605–610.

Silva, JMC Santos and Silvana Tenreyro (2006). "The log of gravity". In: *The Review of Economics and Statistics* 88.4, pp. 641–658.

Staiger, Douglas and James H. Stock (1997). "Instrumental Variables Regression with Weak Instruments". In: *Econometrica* 65.3, pp. 557–586. DOI: 10.2307/2171753.

Tobías, Aurelio et al. (2021). "Geographical variations of the minimum mortality temperature at a global scale: a multicountry study". In: *Environmental Epidemiology* 5.5, e169.

Vrints, Christiaan et al. (2024). "2024 ESC guidelines for the management of chronic coronary syndromes: developed by the task force for the management of chronic coronary syndromes of the European Society of Cardiology (ESC) endorsed by the European Association for Cardio-Thoracic Surgery (EACTS)". In: *European heart journal* 45.36, pp. 3415–3537.

Walters, Christopher (2024). "Empirical Bayes methods in labor economics". In: *Handbook of Labor Economics*. Vol. 5. Elsevier, pp. 183–260.

Yin, Qian et al. (2019). "Mapping the increased minimum mortality temperatures in the context of global climate change". In: *Nature Communications* 10.1, p. 4640.

You, Yanwei et al. (2024). "Inverted U-shaped relationship between sleep duration and phenotypic age in US adults: a population-based study". In: *Scientific reports* 14.1, p. 6247.

Online Appendix

A Data appendix

A.1 GDP

A.1.1 Data

We use an updated version of the panel dataset from Burke et al. (2015), which covers 166 countries observed annually from 1960 to 2019. The outcome variable is annual GDP growth, $\Delta \ln y_{it}$, computed as the first difference of log real GDP. Temperature is measured as the population-weighted annual average in degrees Celsius, constructed from gridded climate reanalysis data (ERA5). The panel is unbalanced; after dropping observations with missing values for $\Delta \ln y_{it}$, temperature, and temperature squared, the estimation sample contains approximately 8,800 country-year observations.

A.1.2 Estimation

We estimate a quadratic temperature–growth relationship in first differences:

$$\Delta \ln y_{it} = \beta_1 T_{it} + \beta_2 T_{it}^2 + \alpha_i + \lambda_t + \gamma_i t + u_{it}, \tag{A1}$$

where T_{it} and T_{it}^2 denote annual average temperature and its square, α_i are country fixed effects, λ_t are year fixed effects, and $\gamma_i t$ are country-specific linear time trends. The implied turning point is $\hat{\tau} = -\hat{\beta}_1/(2\hat{\beta}_2)$.

Our specification differs from Burke et al. (2015) in two respects. First, we do not control for precipitation here (previous studies have found negligible impacts of doing so). Second, we include country-specific linear rather than quadratic time trends.

The pooled model (Figures 1 and 5) estimates common coefficients (β_1, β_2) across all countries. The heterogeneous model (Figure 1) allows fully country-specific (β_{1i}, β_{2i}) by interacting the temperature regressors with country indicators.

A.1.3 Null simulation

The placebo exercise in Figure 1 and Appendix Figure A1 generates data under the null hypothesis that temperature has no effect on GDP growth, while matching the empirical panel structure, residual variation, and temperature process. We proceed in three steps.

Step 1: Calibrate the DGP. We regress $\Delta \ln y_{it}$ on country fixed effects α_i, year fixed effects λ_t, and country-specific linear time trends $\gamma_i t$, omitting all temperature terms. From this regression we extract $\hat{\alpha}_i$, $\hat{\lambda}_t$, and $\hat{\gamma}_i$, as well as the country-specific residual standard deviation $\hat{\sigma}_i = \text{sd}(\hat{u}_{it})$.

Step 2: Simulate under the null. For each country-year observation in the estimation sample, we draw

$$\Delta \ln y_{it}^{\text{sim}} = \hat{\alpha}_i + \hat{\lambda}_t + \hat{\gamma}_i \, t + u_{it}^{\text{sim}}, \quad u_{it}^{\text{sim}} \sim \mathcal{N}(0, \hat{\sigma}_i^2).$$

The simulated outcome preserves cross-country differences in growth levels, global time trends, and country-specific trends, but contains no temperature effect by construction. Temperature variables (T_{it}, T_{it}^2) are held at their empirical values.

Step 3: Estimate. We fit the country-level model of equation (A1) to the simulated data and compute $\hat{\tau}_i^{\text{null}} = -\hat{\beta}_{1i}/(2\hat{\beta}_{2i})$ for each country. The resulting correlation between $\hat{\tau}_i^{\text{null}}$ and mean temperature is nearly as large as in the real data, demonstrating that the adaptation-like pattern arises mechanically from weak identification.

For plotting purposes, we do not include estimates from two countries for which the estimate of τ is way outside the support of the data ($|\hat{\tau}| > 50$).

A.2 Mortality

A.2.1 Data

We use the global mortality panel of Carleton et al. (2022), which links subnational death rates to temperature and precipitation across multiple countries. The unit of observation is an admin-2 region (e.g., county or district) observed annually. Temperature and precipitation are measured using the Global Meteorological Forcing Dataset (GMFD) and enter the regression as population-weighted annual averages. Death rates are winsorized at the 99th percentile within each country–age-group cell to limit the influence of outliers. We use the publicly available sample, which does not include all countries that were used in the original article.

We restrict the sample to the 65+ age group (`agegroup` = 3) since this is the sample for which impacts are most acute. Following the original paper, we restrict to years through 2010. The estimation is population-weighted, with weights normalized within each year.

A.2.2 Estimation

We estimate a quadratic temperature–mortality relationship for the old age group (> 65 yrs):

$$\text{deathrate}_{it} = \sum_k \beta_k \tilde{T}_{it}^k + \mathbf{X}_{it}'\boldsymbol{\delta} + \alpha_i + \lambda_{ct} + u_{it}, \tag{A2}$$

where deathrate_{it}, is the death rate for > 65 year olds in location i in year t. $\tilde{T}_{it}^k = \sum_{d \in t} T_{itd}^k$ is the annual sum of daily temperatures raised to the power k. $\mathbf{X}_{it}$ includes country-specific precipitation controls (linear and quadratic), α_i are admin-2 $\times$ age-group fixed effects, and λ_{ct} are country $\times$ year $\times$ age-group fixed effects.

The pooled model (Figure 5) estimates common (β_1, β_2) across all regions, with population weights. The admin-1-level heterogeneous model (Appendix Figure A12) allows (β_{1j}, β_{2j}) to vary across

admin-1 units.

B Additional figures and tables

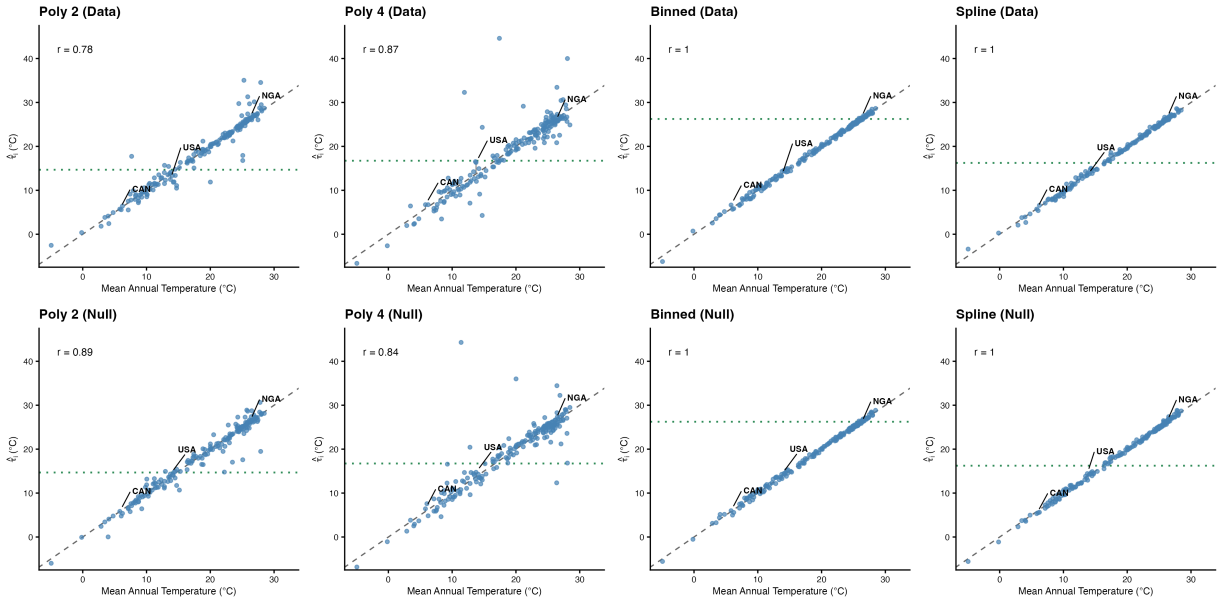


Figure A1: **Spurious adaptation persists across functional-form specifications.** Each column estimates country-specific turning points $\hat{\tau}_i$ using a different specification of the temperature–GDP response function: quadratic (Poly 2), quartic polynomial (Poly 4), 8-bin indicator model (Binned), and natural spline with 2 interior knots (Spline). The top row uses actual GDP panel data; the bottom row uses a placebo simulation in which temperature has no effect on GDP. Points are colored by whether the turning point is a maximum (blue) or minimum (orange). The dashed line is the 45-degree line ($\hat{\tau}_i = \mu_i$); the dotted horizontal line is the pooled optimum. Correlation coefficients (r) between $\hat{\tau}_i$ and mean temperature are reported in each panel. The spurious correlation between estimated optima and local climate persists regardless of specification, confirming that the pattern is driven by weak identification rather than functional-form choice.

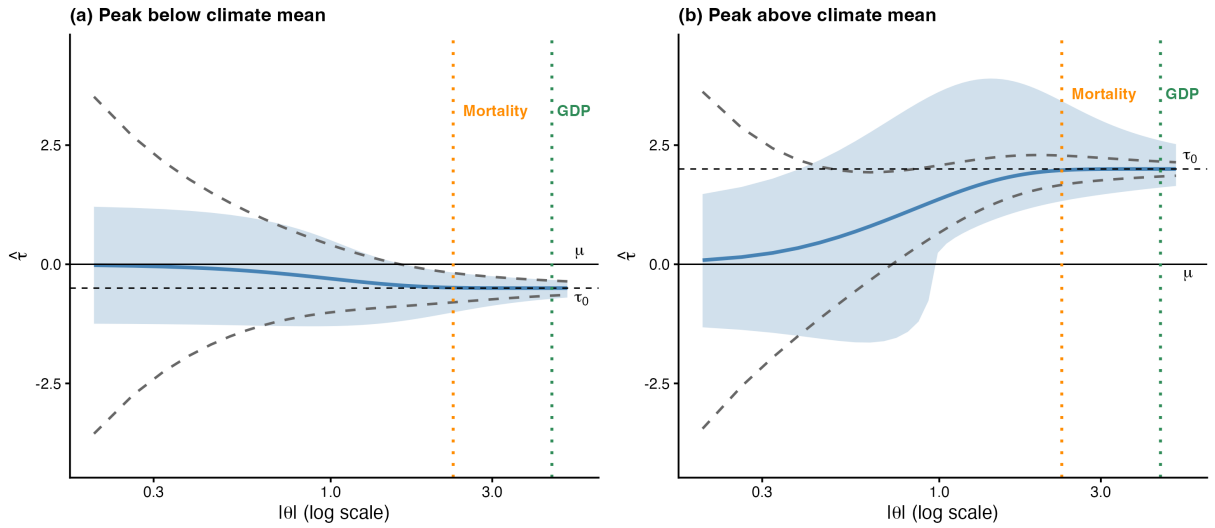


Figure A2: **Median bias and confidence-interval width as a function of curvature strength.** Each panel plots the Hinkley median (solid blue line) and 68% quantile interval (blue band, 16th–84th percentiles) of the turning-point estimator $\hat{\tau}$ as a function of $|\theta|$ (log scale). Grey dashed lines show the corresponding 68% delta-method confidence interval, centered at the Hinkley median. Panel (a): $\tau_0 = \mu - \sigma_x/2$ (peak below climate mean). Panel (b): $\tau_0 = \mu + 2\sigma_x$ (peak above climate mean). Horizontal reference lines mark τ_0 (dashed) and μ (solid). Vertical dotted lines mark the empirical $|\theta|$ for the GDP (green) and mortality (orange) applications. The Hinkley median transitions smoothly from μ (under weak curvature) to τ_0 (under strong curvature), and the delta-method interval diverges from the Hinkley interval when $|\theta|$ is small. Parameters: $\mu = 0$, $\sigma_x = 1$.

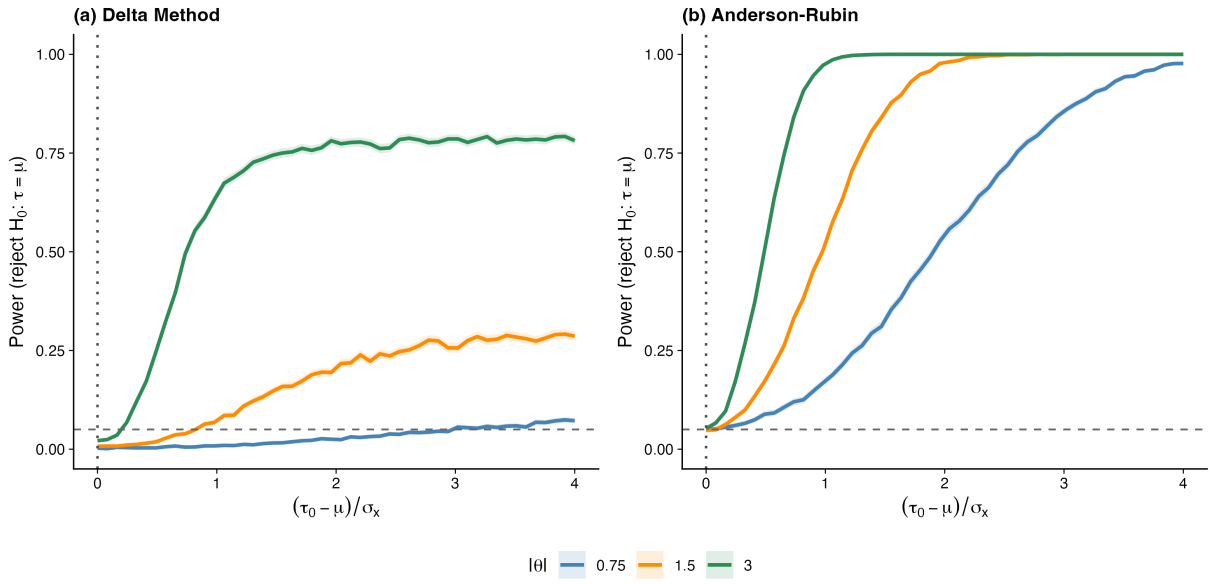
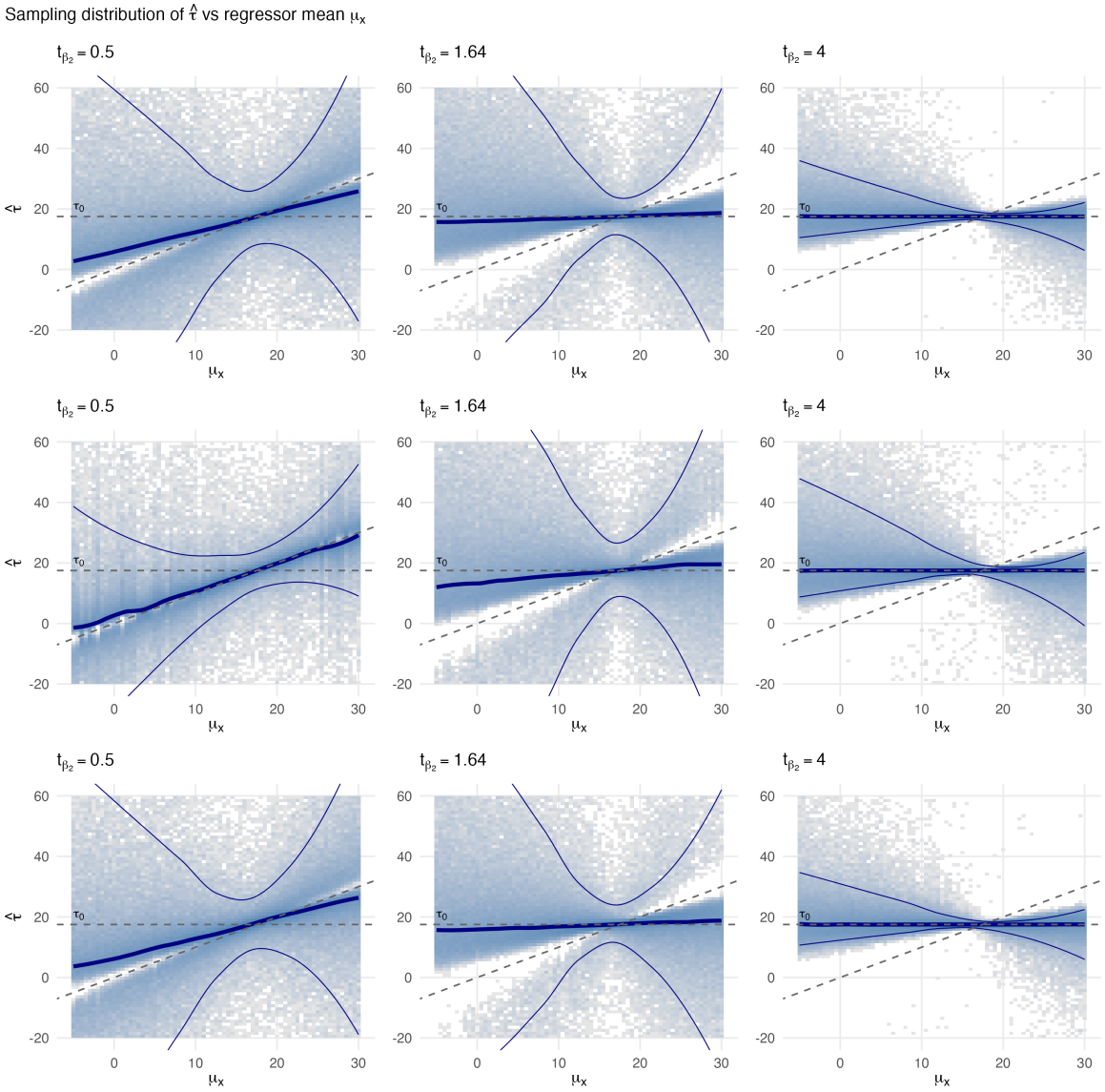


Figure A3: **Power of delta-method and Anderson–Rubin tests for $H_0 : \tau = \mu$.**
Each panel plots rejection rates (power) as a function of the true $(\tau_0 - \mu)/\sigma_x$ for three values of curvature strength: $|\theta| = 0.75$ (blue, weak), $|\theta| = 1.5$ (orange, borderline), and $|\theta| = 3$ (green, strong). Panel (a): delta-method test. Panel (b): Anderson–Rubin (AR) test. Shaded bands show 95% simulation confidence intervals around the estimated power curves. The dashed horizontal line marks the nominal 5% significance level. At $(\tau_0 - \mu)/\sigma_x = 0$ (the null), the AR test is correctly sized across all θ values, while the delta method is severely undersized. Under the alternative, the AR test has substantially higher power, particularly for weak and borderline curvature. $n = 100$; 5,000 simulation replications.



Notes: Shading shows kernel density of simulation draws (n = 100). Solid line: median. Dashed lines: 5th/95th percentiles. Horizontal dashed: true tau. Diagonal dashed: 45-degree line.

Figure A4: **Sampling distribution of $\hat{\tau}$ under alternative data-generating processes.** Each panel shows bootstrap samples of $\hat{\tau}$ against the regressor mean μ_x (transparent points), the bootstrap median (dark blue line), and 95% confidence band (thin blue lines). Columns vary the target curvature strength $|t_{\beta_2}|$; rows vary the DGP. Top row: baseline DGP from Proposition 2.1 (Gaussian x, homoskedastic errors). Middle row: heteroskedastic errors with $\text{Var}(u_i \mid x_i) = \sigma_u^2(1 + ((x - \mu)/\sigma_x)^2)^2$. Bottom row: x_i drawn from a skewed normal distribution, holding the mean and variance fixed at μ and σ_x^2. The dashed diagonal line is the 45-degree line ($\hat{\tau} = \mu_x$); the horizontal dashed line marks the true τ_0. The qualitative pathologies of Proposition 2.1—median bias toward μ, fat tails, and bimodality under weak curvature—are robust to heteroskedasticity and non-normal regressors. $n = 100$; $\sigma_x = 2$.

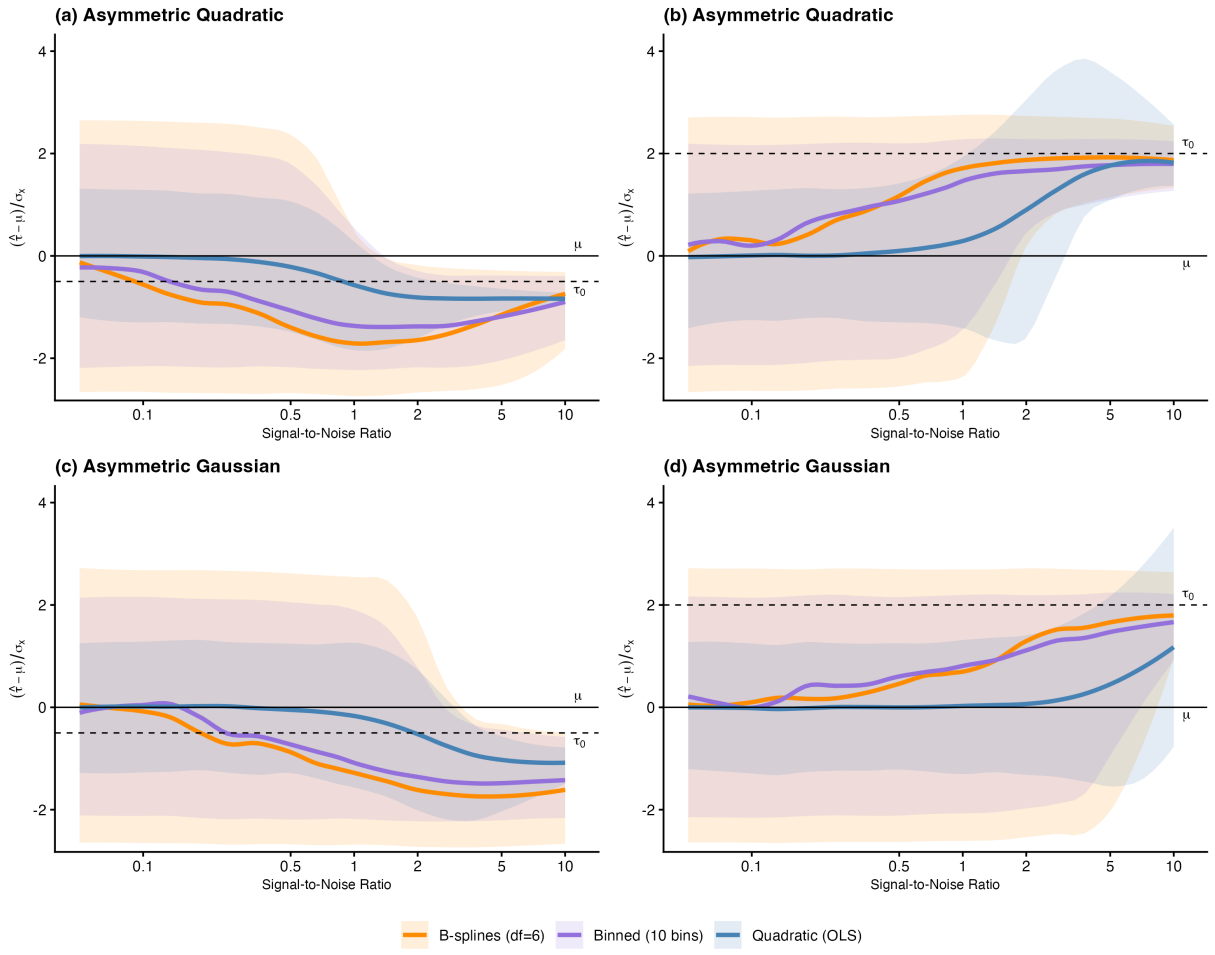


Figure A5: **Weak identification persists under misspecified DGPs.**
Comparison of quadratic (blue), cubic B-spline with 6 d.f (orange), and binned (10 bins, purple) estimators for $\hat{\tau}$ under two non-quadratic response functions. Top row: asymmetric piecewise-quadratic (gentle cold side, steep hot side). Bottom row: asymmetric Gaussian (non-polynomial exponential peak with different spreads on each side). Left column: $\tau_0 = \mu - \sigma_x/2$ (peak near climate). Right column: $\tau_0 = \mu + 2\sigma_x$ (peak far from climate). Lines show the median of $(\hat{\tau} - \mu)/\sigma_x$; shaded bands show 68% Monte Carlo confidence intervals. All three estimators converge toward μ as the signal-to-noise ratio deteriorates, confirming that the weak-identification problem is fundamental rather than an artifact of functional-form assumptions. The semi-parametric estimators exhibit less median bias but substantially higher variance than OLS when the quadratic model is misspecified. $n = 200$; 2,000 replications; $\sigma_x \approx 2$.

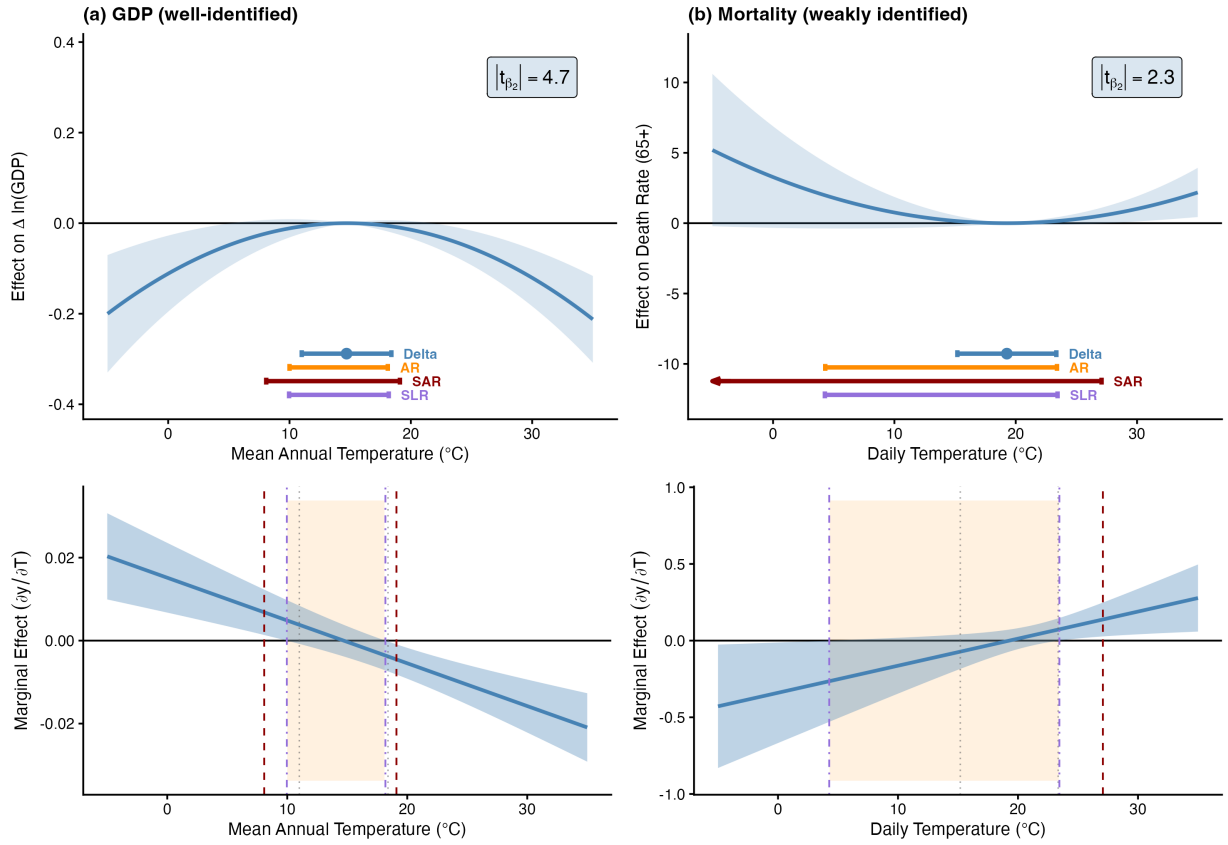


Figure A6: **Delta-method versus weak identification-robust inference on turning points.** *Top row.* Estimated response functions for GDP growth (left, Burke et al., 2015) and 65+ age-group mortality rates (right, Carleton et al., 2022), with pointwise 95% confidence bands. Horizontal bars report four 95% confidence sets for the turning point τ_0: the delta-method interval (blue); the Anderson–Rubin interval (AR, orange), which inverts a χ^2_1 test of $H_0\colon \tau_0 = \tau$; the sign-restricted Anderson–Rubin interval (SAR, dark red); and the sign-restricted likelihood-ratio interval (SLR, purple). SAR and SLR additionally impose the maintained assumption that the response is concave (GDP) or convex (mortality). *Bottom row.* AR inversion: for each hypothesized τ_0, the marginal effect $\hat{\beta}_1 + 2\hat{\beta}_2\tau_0$ and its delta-method confidence band (steel blue ribbon) are plotted; the AR confidence set (orange shading) collects all τ_0 for which this band includes zero. Vertical lines mark the delta-method point estimate (grey dotted), SAR bounds (dark red dashed), and SLR bounds (purple dot-dash). Further details on data and estimation in Appendix A.

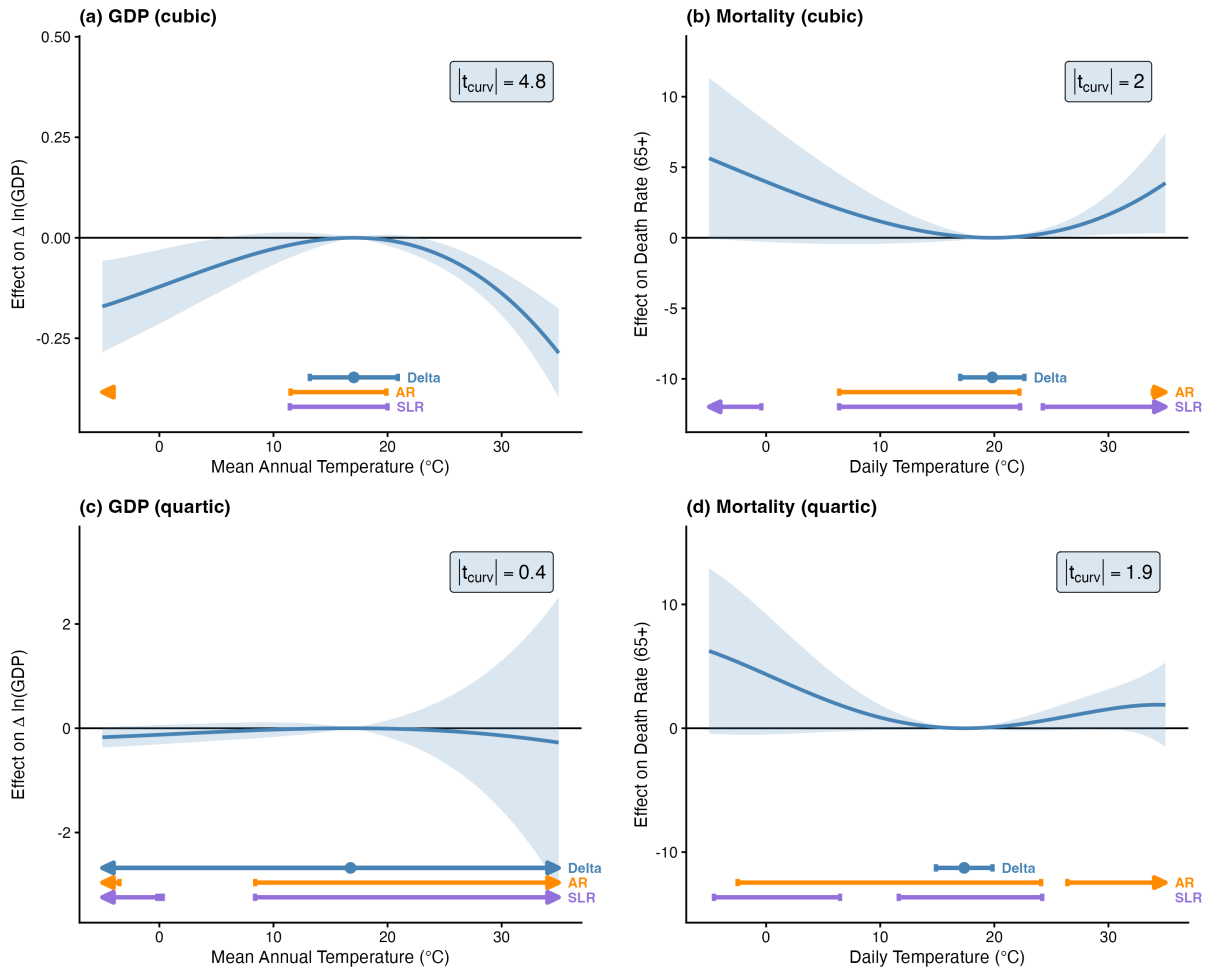


Figure A7: **Delta-method versus weak identification-robust inference on turning points: cubic and quartic specifications.**
Left column: GDP growth. Right column: 65+ mortality rates. Top row: cubic. Bottom row: quartic. Each panel plots the estimated response function with pointwise 95% confidence bands, with horizontal bars reporting three 95% confidence sets for the turning point τ_0: the delta-method interval (blue); the Anderson–Rubin interval (AR, orange), which inverts a χ^2_1 test of $H_0\colon f'(\tau_0) = 0$; and the sign-restricted likelihood-ratio interval (SLR, purple), which additionally imposes the second-order condition $f''(\tau_0) < 0$ (GDP) or $f''(\tau_0) > 0$ (mortality). Arrows indicate confidence sets that extend beyond the plotted range. The AR and SLR sets generalize the quadratic constructions in Figure 5 by testing the first-order condition of a degree-K polynomial; the SLR set imposes the second-order condition as a sign restriction on the curvature $\mathbf{c}_K'\boldsymbol{\beta}$ at each hypothesized τ_0. Curvature t-statistics are reported in each panel. Data and fixed-effect specifications follow Appendix A.

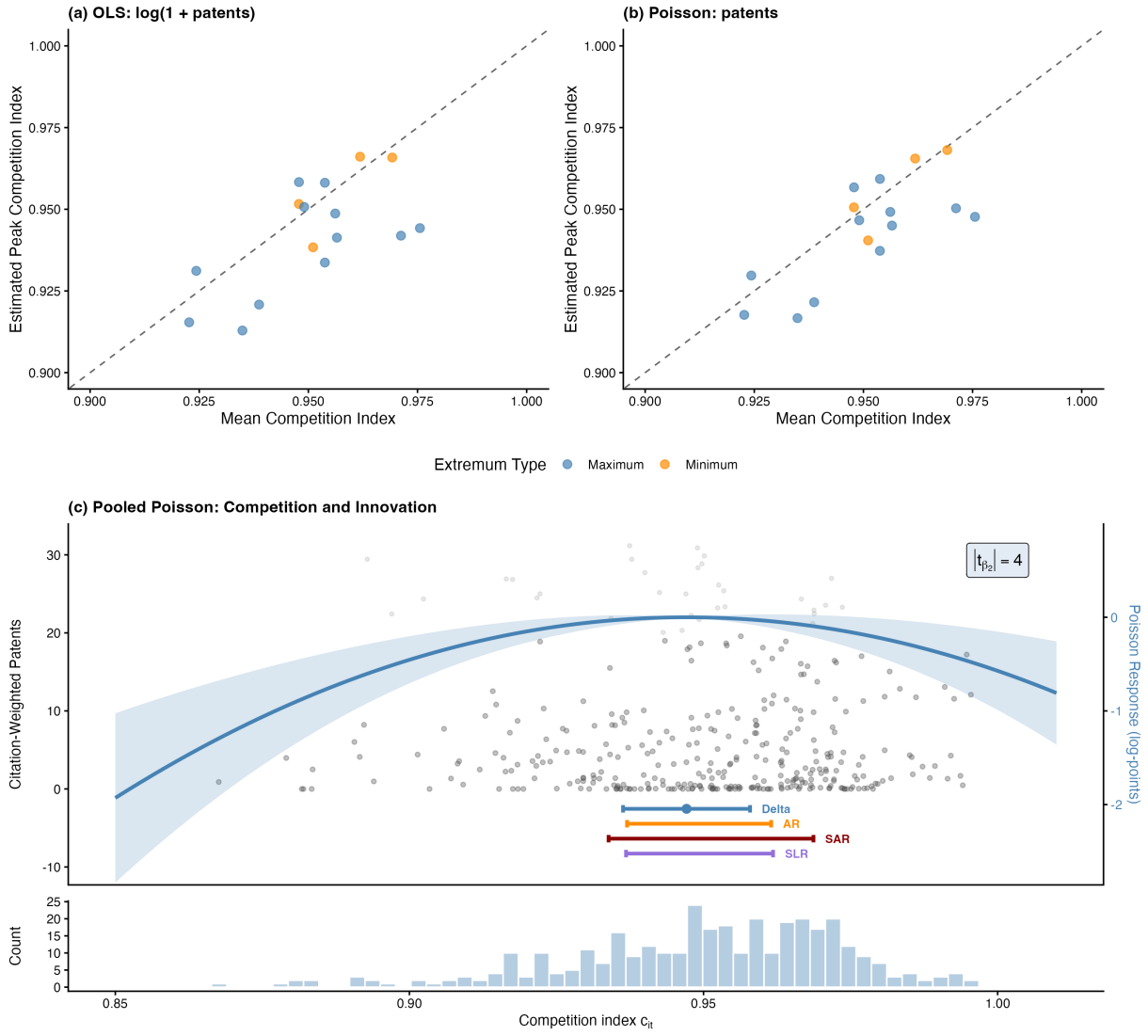


Figure A8: **Competition and innovation: industry-level peaks and pooled response function.** Data from Aghion et al. (2005) (UK manufacturing industries, 1973-1994). The outcome is citation-weighted patents; the regressor of interest is the Lerner competition index c_{it} for industry i in year t. Panels (a) and (b) plot industry-level estimates of the peak competition index $\hat{c}_i^*$ against mean industry i competition $\bar{c}_i$, estimated by OLS on $\log(1 + \text{patents}_{it})$ and by Poisson pseudo-maximum likelihood, respectively. Blue: industries where $\hat{\beta}_{2i} < 0$ (maximum); orange: $\hat{\beta}_{2i} > 0$ (minimum). Dashed line: 45-degree benchmark. Panel (c) shows the pooled Poisson response function with industry and year fixed effects and HC1 standard errors. Grey points are the raw data (darker points lie between the 10th and 90th percentiles of the patent distribution, replicating the data in Figure 1 of Aghion et al., 2005). The blue curve and shaded region show the estimated response function and 95% confidence band. Horizontal bars at the bottom display the delta-method and Anderson–Rubin 95% confidence intervals for the peak c^*; the AR interval is 1.1× wider.

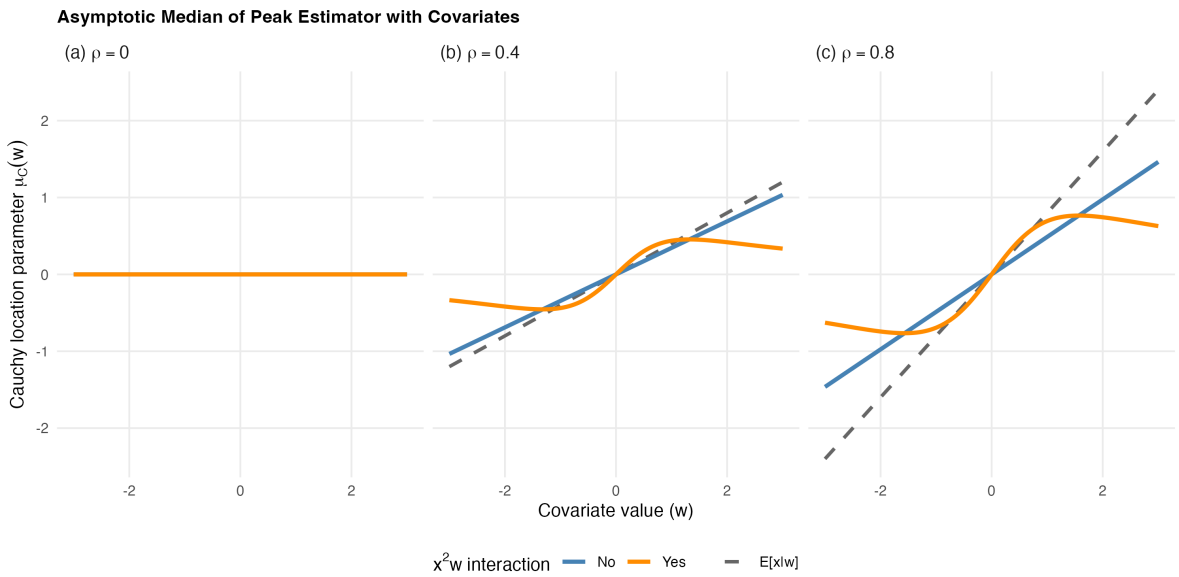


Figure A9: **Cauchy median of $\hat{\tau}(w)$ under covariate-augmented models.**
Each panel plots the Cauchy location parameter $\check{\mu}(w)$—the asymptotic median of $\hat{\tau}(w)$ under the null of no temperature effect—for models in which both the peak τ_i and curvature γ_i are linear functions of w_i. Panels vary the correlation ρ between x_i and w_i. Blue: model without $x_i^2 w_i$ interaction; orange: model including the $x_i^2 w_i$ interaction. Grey dashed line: $\mathbb{E}[x \mid w]$ for jointly normal (x, w). When $\rho = 0$, $\check{\mu}(w) = \mu$ regardless of the model. When $\rho > 0$, the median bias tracks $\mathbb{E}[x \mid w]$ over most of the support, especially when the nonlinear interaction is included. Both x_i and w_i are standardized to zero mean and unit variance.

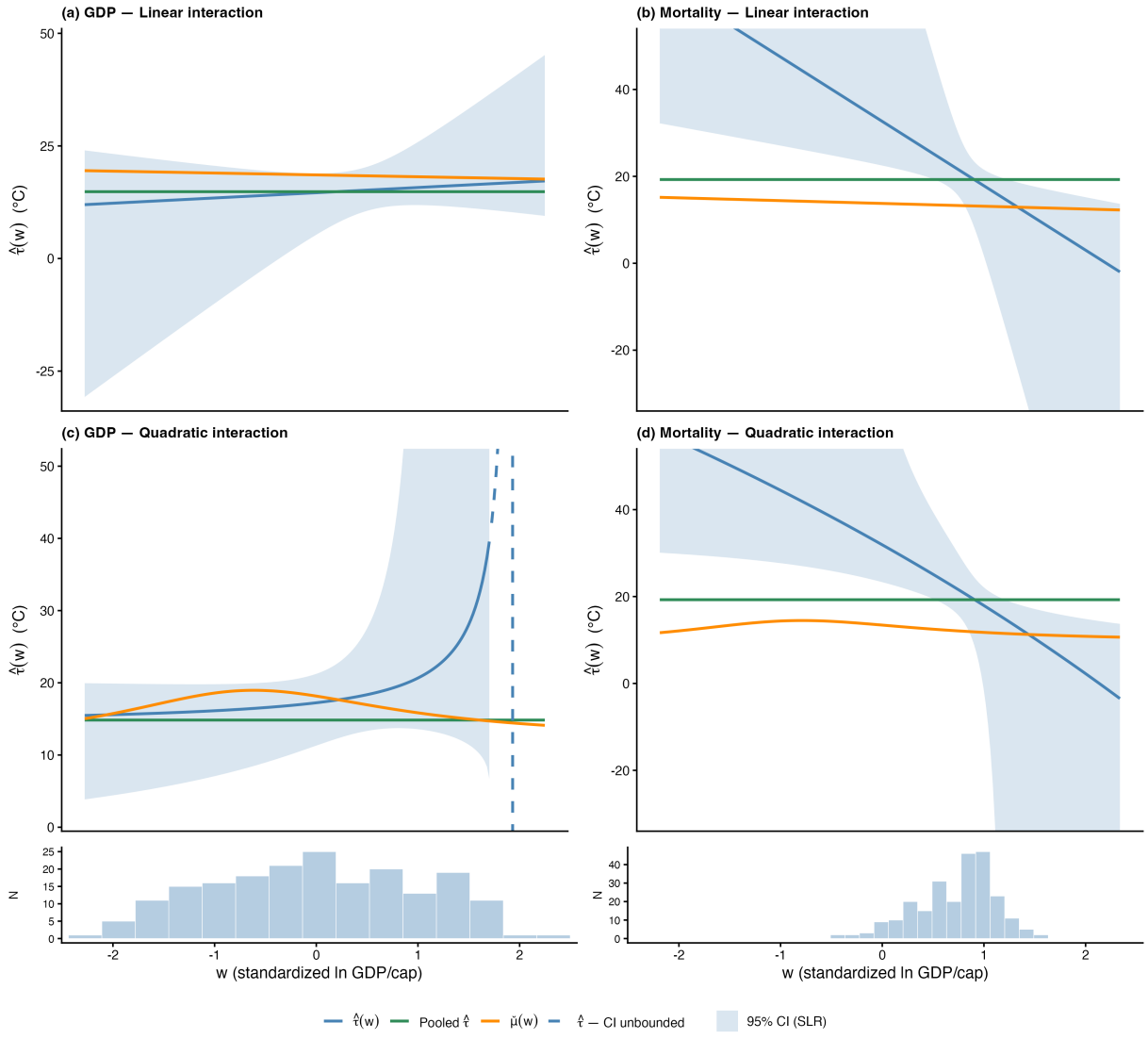


Figure A10: **Covariate heterogeneity in peak temperature estimates.**
Each panel plots $\hat{\tau}(w)$, the estimated peak temperature as a function of standardized log GDP per capita w, with 95% sign-restricted likelihood ratio (SLR) confidence bands, using the estimates reported in Table A3. The top row uses a linear interaction ($T \times w$ only); the bottom row adds a quadratic interaction ($T^2 \times w$). Left column: GDP growth; right column: mortality. The solid green line shows the pooled $\hat{\tau}$ (no interaction). The solid orange line shows $\check{\mu}(w)$, the Cauchy location of $\hat{\tau}(w)$ under the null of no effect (Corollary 3.1). Dashed segments of $\hat{\tau}(w)$ indicate regions where the SLR confidence set is unbounded. Histograms at the bottom show the distribution of w across units.

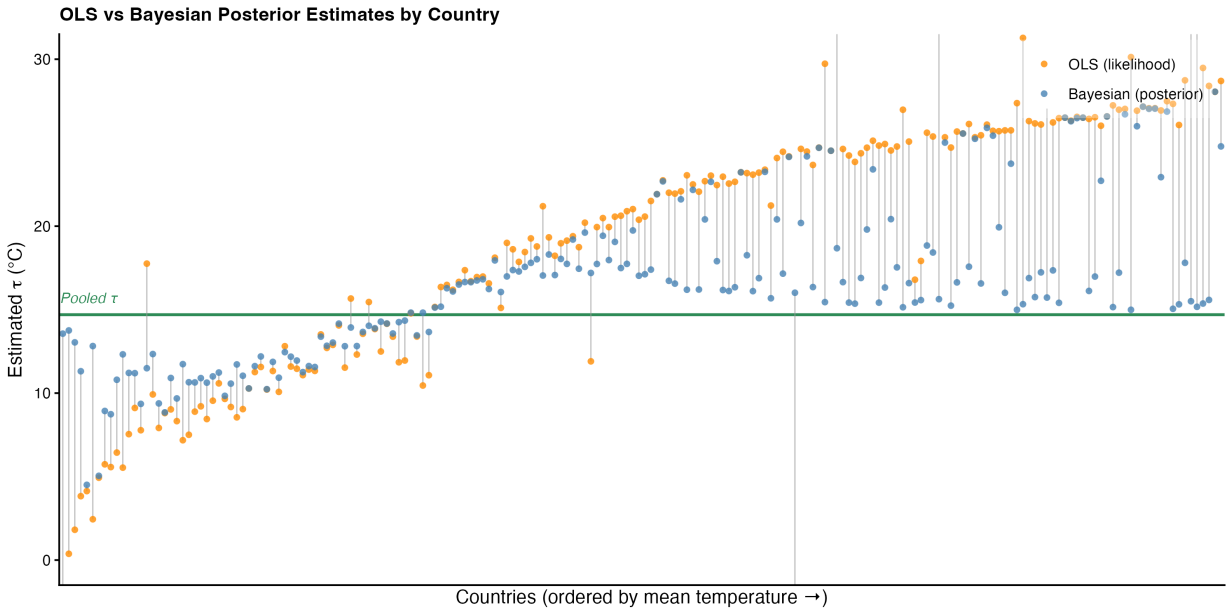


Figure A11: **Country-level shrinkage from OLS to Bayesian posterior in the temperature–GDP sample.**
OLS and Bayesian posterior estimates of $\hat{\tau}_i$ for each country, ordered by mean temperature (coldest at left). Orange points are OLS estimates ($-\frac{1}{2}\hat{\beta}_{1i}/\hat{\beta}_{2i}$); blue points are posterior means under the pooled prior described in Section 3.2. Grey segments connect each country's OLS and Bayesian estimates, visualizing the direction and magnitude of shrinkage. The solid green line marks the pooled $\hat{\tau}$. Triangles indicate estimates clipped to the displayed $(0, 30)$°C range. The posterior pulls noisy OLS estimates—particularly those far from the pooled value—toward the common prior, while countries with precisely estimated curvature retain OLS-like peaks.

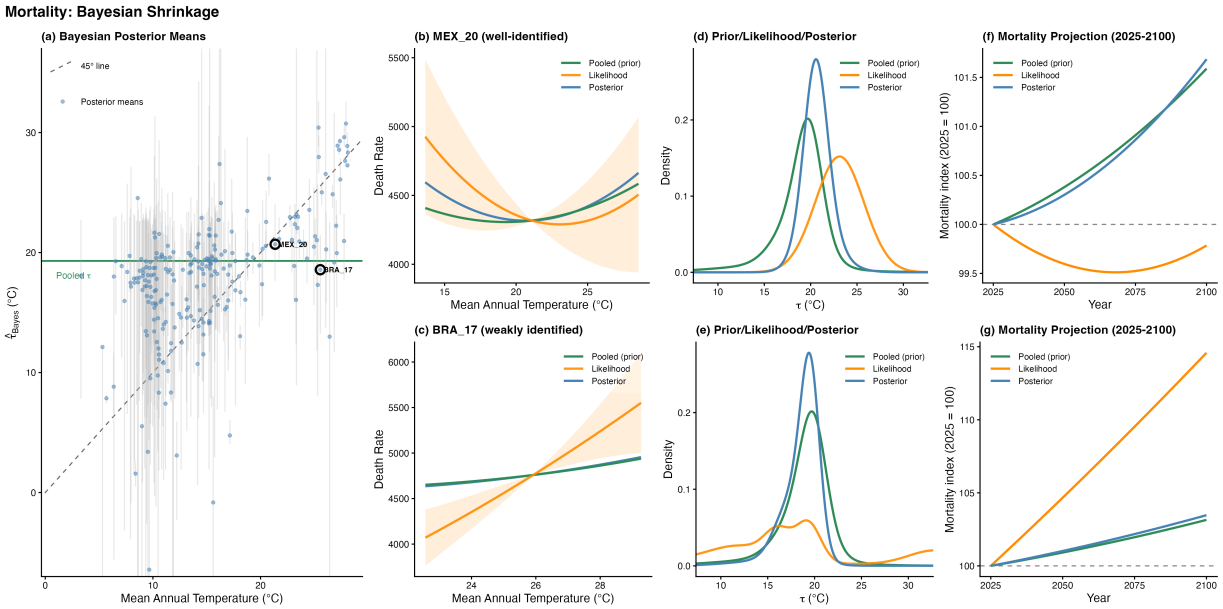


Figure A12: **Bayesian shrinkage toward the pooled estimate in the temperature–mortality relationship.**

Mortality (65+ age group), estimated at the admin-1 level. Panel (a): posterior mean $\hat{\tau}_{\text{Bayes}}$ against admin-1 mean temperature, with 95% posterior credible intervals (grey bars). The dashed line is the 45-degree line ($\hat{\tau} = \bar{x}$); the solid green line is the pooled $\hat{\tau}$. Circled points are the example admin-1 units shown in panels (b)–(g). Panels (b)–(c): scatter of death rates against temperature with pooled (green) and admin-1-specific (orange) quadratic fits and 95% confidence bands for a well-identified and a weakly identified unit. Panels (d)–(e): prior (pooled sampling distribution, green), likelihood (admin-1-specific Hinkley likelihood, orange), and posterior (blue) densities for the same two units. Panels (f)–(g): projected mortality index (2025 = 100) under uniform $+3$°C warming from 2025–2100, comparing posterior (blue) and pooled (green) damage estimates.

B.1 Tables

	Relative median bias		AR width excess		SLR width excess	
$\lvert t_{\beta_2}\rvert$	$\tau_0 = \mu - \frac{\sigma_x}{2}$	$\tau_0 = \mu + 2\sigma_x$	$\tau_0 = \mu - \frac{\sigma_x}{2}$	$\tau_0 = \mu + 2\sigma_x$	$\tau_0 = \mu - \frac{\sigma_x}{2}$	$\tau_0 = \mu + 2\sigma_x$
Panel A: Analytic						
0.5	192%	45%	∞	∞	∞	∞
1	86%	20%	∞	∞	∞	∞
1.3	46%	11%	∞	∞	∞	∞
1.64	20%	5%	∞	∞	∞	∞
1.96	9%	2%	∞	∞	∞	∞
2.17	5%	1%	86%	86%	86%	86%
2.5	2%	<1%	73%	71%	73%	71%
$\sqrt{10}$	<1%	<1%	48%	47%	48%	47%
4	<1%	<1%	28%	28%	28%	28%
5	<1%	<1%	17%	16%	17%	16%
7.5	<1%	<1%	7%	7%	7%	7%
10	<1%	<1%	4%	4%	4%	4%
12.5	<1%	<1%	2%	2%	2%	2%
Panel B: Bootstrap (heteroskedastic, skewed x)						
$[0,\ 0.5)$	80%	90%	∞	∞	∞	∞
$[0.5,\ 1)$	98%	99%	∞	∞	∞	∞
$[1,\ 1.3)$	114%	99%	∞	∞	∞	∞
$[1.3,\ 1.64)$	97%	92%	∞	∞	∞	∞
$[1.64,\ 1.96)$	54%	43%	∞	∞	>999%	600%
$[1.96,\ 2.17)$	37%	28%	548%	909%	546%	513%
$[2.17,\ 2.5)$	27%	21%	156%	229%	160%	229%
$[2.5,\ \sqrt{10})$	29%	27%	64%	89%	65%	89%
$[\sqrt{10},\ 4)$	17%	13%	30%	41%	31%	41%
$[4,\ 5)$	11%	9%	17%	23%	18%	23%
$[5,\ 7.5)$	9%	12%	9%	11%	9%	11%
$[7.5,\ 10)$	5%	5%	4%	5%	4%	5%
$[10,\ 12.5)$	3%	2%	2%	3%	2%	3%

Table A1: **Diagnostics for a quadratic peak estimator.**
The true DGP is $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ with $\tau^* = -\beta_1/(2\beta_2)$. Panel A reports analytic results: relative median bias is computed from the Hinkley (1969) PDF of the ratio estimator; AR and SLR width excess use 50,000 Monte Carlo draws from the known V matrix. Panel B reports bootstrap results ($n = 100$) with at least 500 accepted draws per bin, binned by realized $|t_{\beta_2}|$; each bin sets curvature $\beta_2 = -\bar{\theta}$ (bin midpoint) with σ_u calibrated so that $E[|t_{\beta_2}|] \approx \bar{\theta}$. The DGP combines heteroskedastic errors, $u_i = \varepsilon_i \cdot (1 + ((x_i - \mu)/\sigma_x)^2)$ with $\varepsilon_i \sim N(0, \sigma_u)$, and skewed regressors $x_i \sim \mathrm{SN}(\alpha = 5)$ calibrated to $E[x] = \mu = 0$ and $\mathrm{SD}(x) = \sigma_x = 2$. Standard errors use the HC (sandwich) estimator. Relative median bias is $|\mathrm{median}(\hat{\tau}) - \tau^*|/|\mu - \tau^*|$. Width excess is $(\mathrm{width}/\mathrm{width}_\Delta - 1) \times 100\%$, median among draws with bounded sets. SLR imposes $\beta_2 < 0$ (maximum).

	Relative median bias		AR width excess		SLR width excess	
$\lvert t_{\text{curv}}\rvert$	$\tau_0 = \mu - \frac{\sigma_x}{2}$	$\tau_0 = \mu + 2\sigma_x$	$\tau_0 = \mu - \frac{\sigma_x}{2}$	$\tau_0 = \mu + 2\sigma_x$	$\tau_0 = \mu - \frac{\sigma_x}{2}$	$\tau_0 = \mu + 2\sigma_x$
Panel A: Cubic estimator						
[0, 0.5)	52%	82%	∞	∞	∞	∞
[0.5, 1)	89%	91%	∞	∞	∞	∞
[1, 1.3)	50%	53%	∞	>999%	∞	187%
[1.3, 1.64)	21%	44%	∞	>999%	∞	104%
[1.64, 1.96)	21%	39%	>999%	>999%	46%	109%
[1.96, 2.17)	19%	38%	>999%	>999%	27%	135%
[2.17, 2.5)	15%	37%	>999%	>999%	23%	143%
[2.5, $\sqrt{10}$)	10%	33%	>999%	>999%	13%	83%
[$\sqrt{10}$, 4)	7%	28%	>999%	>999%	8%	36%
[4, 5)	<1%	23%	>999%	>999%	3%	18%
[5, 7.5)	2%	15%	>999%	>999%	1%	11%
[7.5, 10)	2%	7%	>999%	>999%	<1%	5%
[10, 12.5)	<1%	4%	>999%	>999%	<1%	2%
Panel B: Quartic estimator						
[0, 0.5)	85%	68%	∞	∞	∞	∞
[0.5, 1)	79%	90%	∞	∞	∞	∞
[1, 1.3)	16%	42%	∞	>999%	∞	>999%
[1.3, 1.64)	20%	40%	∞	>999%	797%	854%
[1.64, 1.96)	9%	28%	>999%	>999%	726%	868%
[1.96, 2.17)	3%	24%	>999%	>999%	686%	753%
[2.17, 2.5)	<1%	21%	>999%	>999%	254%	236%
[2.5, $\sqrt{10}$)	1%	22%	>999%	>999%	92%	234%
[$\sqrt{10}$, 4)	<1%	16%	>999%	>999%	15%	27%
[4, 5)	2%	11%	>999%	>999%	7%	11%
[5, 7.5)	2%	5%	>999%	>999%	2%	4%
[7.5, 10)	1%	2%	>999%	>999%	<1%	<1%
[10, 12.5)	<1%	2%	>999%	>999%	<1%	<1%

Table A2: **Diagnostics for cubic and quartic peak estimators.**
The true DGP is $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ with $\beta_2 = -1$; the estimated model is a degree-p polynomial ($p = 3$ cubic, $p = 4$ quartic). Regressors $x_i \sim N(0, 2)$, errors $u_i \sim N(0, \sigma_u)$, $n = 100$. Each bin calibrates σ_u so that $E[|t_{\text{curv}}|] \approx \bar{\theta}$ (bin midpoint), with at least 500 accepted draws per bin. The curvature t-statistic is $|t_{\text{curv}}| = |f''(\hat{\tau})|/\text{se}(f''(\hat{\tau}))$ where $f''(x) = \sum_{j=2}^{p} j(j-1)\hat{\beta}_j x^{j-2}$. Turning point $\hat{\tau}$ is the root of $f'(x) = 0$ with $f''(\hat{\tau}) < 0$ (global maximum). Relative median bias is $|\text{median}(\hat{\tau}) - \tau^*|/|\mu - \tau^*|$. AR inverts the Wald test for $f'(\tau_0) = 0$ on a grid; SLR additionally constrains $f''(\tau_0) \leq 0$ (concavity at the turning point). Width excess is median$(w_{\text{AR}}/w_{\Delta} - 1)$ or median$(w_{\text{SLR}}/w_{\Delta} - 1)$ among draws with bounded confidence sets. When confidence sets consist of disjoint intervals, their total width is the sum of individual interval widths.

	GDP Growth			Mortality		
	No-w	Linear	Quadratic	No-w	Linear	Quadratic
(A) Coefficient Estimates						
T	0.0091	0.0083	0.0123	-0.3401	-0.6158	-0.6291
	(0.0026)	(0.0037)	(0.0042)	(0.1672)	(0.1827)	(0.1961)
$T \times w$	—	0.0007	-0.0049	—	0.2803	0.2958
		(0.0015)	(0.0030)		(0.0931)	(0.1738)
T^2	-0.0003	-0.0003	-0.0004	0.0088	0.0096	0.0100
	(0.0001)	(0.0001)	(0.0001)	(0.0039)	(0.0039)	(0.0040)
$T^2 \times w$	—	—	0.0002	—	—	-0.0005
			(0.0001)			(0.0037)
(B) Fixed Effects and Controls						
Unit FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Unit $\times$ Year Trend	✓	✓	✓			
Unit $\times$ Precip.				✓	✓	✓
Unit $\times$ Year $\times$ Age				✓	✓	✓
Observations	8,805	8,786	8,786	137,078	137,078	137,078
Adj. R^2	0.138	0.138	0.139	0.764	0.764	0.764
(C) Peak Temperature (at $w = 0$)						
$\hat{\tau}$	14.8	14.6	17.2	19.3	32.0	31.3
Delta CI (95%)	$[11.2, 18.5]$	$[9.8, 19.3]$	$[13.1, 21.3]$	$[15.2, 23.3]$	$[18.4, 45.6]$	$[20.5, 42.1]$
AR CI (95%)	$[10.6, 18.9]$	$[4.7, 18.8]$	$[11.1, 21.0]$	$[4.3, 23.4]$	$[22.3, 80.0]$	$[23.1, 66.6]$
$\lvert t_{\beta_2} \rvert$	3.97	2.84	3.64	2.29	2.44	2.53

Table A3: **Quadratic temperature regressions with covariate interactions.**
Panel (A) reports coefficient estimates from quadratic temperature regressions for GDP growth (columns 1–3) and mortality (columns 4–6). T is temperature in degrees Celsius and w is standardized (mean zero, unit variance) average log GDP per capita in constant 2015 dollars from the World Development Indicators. "No-w": pooled baseline with no covariate interaction. "Linear": w interacted with T only. "Quadratic": w interacted with both T and T^2. Panel (B) lists fixed effects and controls: the GDP specification includes unit, year, and unit-specific linear time trends; the mortality specification includes unit, year, unit-specific precipitation controls, and unit $\times$ year $\times$ age-group effects. Panel (C) reports the implied peak temperature $\hat{\tau}$ evaluated at $w = 0$, with both conventional delta-method and identification-robust Anderson–Rubin 95% confidence intervals, and the absolute t-statistic on the curvature term $|t_{\beta_2}|$. Standard errors in parentheses, clustered at the unit level.

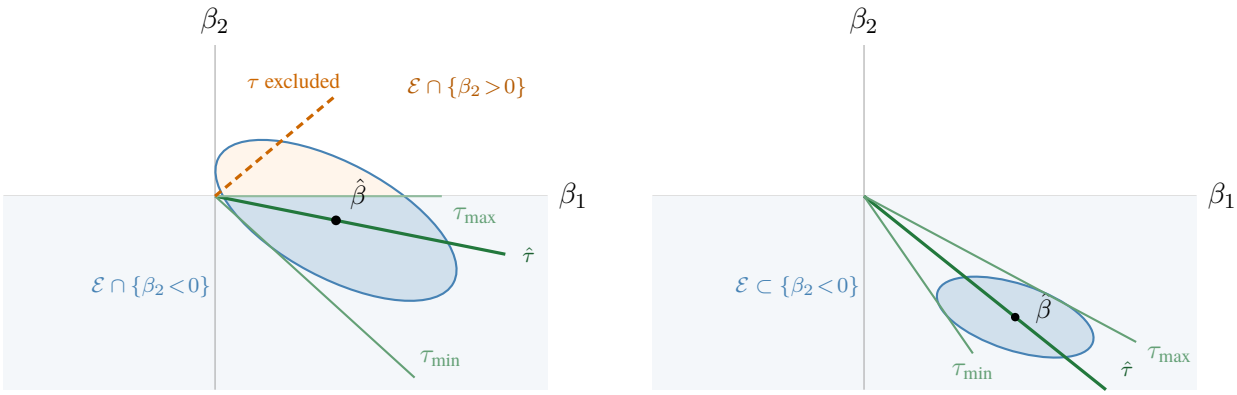


Figure A13: Geometric intuition for the sign-restricted Anderson-Rubin (SAR) confidence set. Green rays from the origin trace values of $\tau = -\beta_1/(2\beta_2)$ through the joint confidence ellipse $\mathcal{E}_2$; the darkest ray marks $\hat{\tau}$, while τ_{min} and τ_{max} bound the set. Orange dashed rays correspond to τ values excluded by the sign restriction $\beta_2 < 0$. *Left:* $\hat{\beta}$ lies near $\beta_2 = 0$, so the ellipse straddles the axis and the confidence set is wide. *Right:* $\hat{\beta}$ is well below $\beta_2 = 0$, so the ellipse lies entirely in $\{\beta_2 < 0\}$ and the confidence set is compact.

C Robust confidence sets: extensions

C.1 Sign-restricted Anderson-Rubin (SAR) confidence sets

Suppose τ_0 reflects a maximum, as in the temperature-GDP relationship (it is straightforward to extend all our theory to the case in which τ_0 is a minimum). Anderson-Rubin confidence sets provide correct coverage for *a* turning point τ_0, but do not impose the nature of the turning point. This can be problematic, as when curvature is weak, Anderson-Rubin confidences sets can be disjoint, covering economically irrelevant cases. Intuitively, when curvature is weak, imposing $\beta_2 < 0$ should rule out some parts of the Anderson-Rubin confidence set, but at the cost of expanding the width of the remaining interval.

To capture this intuition, we work in the space of (β_1, β_2). Letting, $\hat{\beta} \sim \mathcal{N}(\beta, \Sigma)$, we can consider the Wald ellipse for confidence level α,

$$\mathcal{E} = \left\{ \beta : (\hat{\beta} - \beta)' \Sigma^{-1} (\hat{\beta} - \beta) \leq \chi^2_{2,1-\alpha} \right\}.$$

where $\chi^2_{2,1-\alpha}$ is the α-critical value of a chi-squared distribution with two degrees of freedom. We

use two degrees of freedom, as here the confidence set reflects a joint hypothesis on $\hat{\tau}$ and β_2.[14] Any turning point $\tau_0 = -\beta_1/(2\beta_2)$ can be represented in this space as a ray through the origin parameterized by s, $\mathcal{R}(\tau_0) = (-2\tau_0 s, s)$. For $\beta_2 < 0$, this is the ray segment $s < 0$. A candidate τ is included in the confidence set if $R(\tau)$ intersects $\mathcal{E}$, so

$$\mathrm{CS}_{\alpha}^{SAR} = \{\tau : \mathcal{E} \cap \mathcal{R}(\tau) \neq \emptyset\}.$$

Figure A13 provides geometric intuition for this set. The figure on the left represents the construction of an SAR confidence set when β_2 is weakly identified and $\mathcal{E}$ contains values $\beta_2 > 0$. Green rays $\mathcal{R}(\tau)$ are pictured coming out of the origin, intersecting the 'Wald' ellipse $\mathcal{E}$. The dark orange dashed line represents a τ value that value that would be included in a traditional Anderson-Rubin confidence set, but is excluded from our SAR set as it does not correspond to a maximum. In the right panel, β_2 is strongly identified, and the confidence set is not affected by the sign restriction.

Comparison of confidence sets

Figure A14 extends the coverage comparison in Figure 4 to include the sign-restricted Anderson–Rubin (SAR) confidence set. The SAR set is conservative relative to SLR, overcovering when the sign restriction $\beta_2 < 0$ is correctly imposed, but shares the desirable property that coverage goes to zero when the restriction is violated.

[14] Note that at the boundary ($\beta_2 = 0$), the chi-bar distribution provides correct critical values (Self and Liang, 1987).

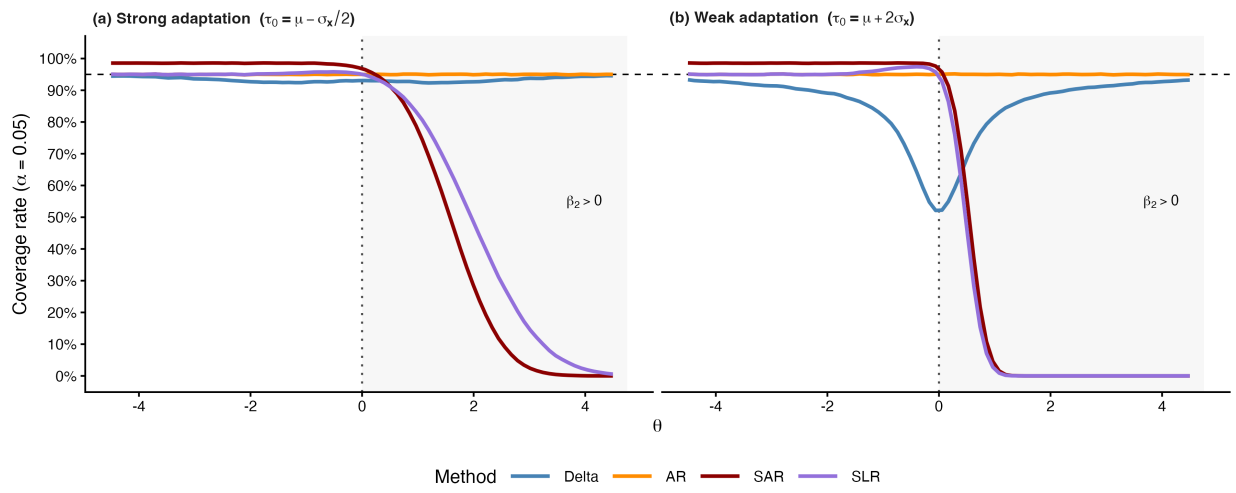


Figure A14: **Coverage of sign-restricted confidence sets, including the projection method.** Each panel plots empirical coverage of nominal 95% confidence sets for the turning point τ_0 as a function of curvature strength θ, under the maintained restriction $\beta_2 < 0$. The horizontal dashed line marks the nominal 95% level; departures below indicate undercoverage. The shaded region ($\theta > 0$) corresponds to $\beta_2 > 0$, where the sign restriction is violated. Left: $\tau_0 = \mu - \sigma_x/2$ (peak near climate). Right: $\tau_0 = \mu + 2\sigma_x$ (peak far from climate). The delta method (orange) under-covers when θ is small and the peak is far from climate. The Anderson–Rubin set (AR, blue) achieves correct coverage throughout. The sign-restricted projection set (SAR, maroon) is conservative, over-covering when $\theta < 0$. The sign-restricted likelihood ratio set (SLR, lilac) matches AR when $\beta_2 < 0$ and is less conservative than SAR. Coverage for both sign-restricted tests goes to zero as θ grows positive, consistent with correctly rejecting turning points that are not maxima. Based on 20,000 Monte Carlo draws from the DGP of Proposition 2.1.

Figure A6 reproduces the applications in Figure 5, adding the sign-restricted Anderson–Rubin (SAR) confidence set. The SAR set is wider than the SLR set in both applications, reflecting the conservatism of the projection approach. When curvature is strongly identified (GDP, left), the difference is modest. When curvature is weaker (mortality, right), the SAR set is substantially wider than both the AR and SLR sets.

C.2 Extension to Poisson estimators

In many contexts, the left-hand side variable is given in logs. If the conditional mean is misspecified; e.g., exogeneity between x_i and u_i is given by $\mathbb{E}[\exp(u_i) \mid x_i] = 1$, rather than $\mathbb{E}[u_i \mid x_i] = 0$, then taking logs and running OLS can generate biased estimates for β_1 and β_2 and Poisson regression

may be more appropriate (Silva and Tenreyro, 2006). We examine whether this bias can affect estimates of the innovation-maximising level of competition across U.K. manufacturing industries using data from Aghion et al. (2005). Appendix Figure A8 estimates industry-level turning points for the effect of a competition index on citation weighted patents, in both a log-plus-one OLS model and a Poisson model. Estimated turning points are near average levels of industry competitiveness for both models. A naive reading of the data suggests that level of competition is at its innovation-maximising level across manufacturing industries in the U.K.. The pooled peak is well-identified, $|\hat{t}_{\beta_2}| \approx 4$, and its Anderson-Rubin confidence set is about 10% wider than that implied by the delta-method.

D Proofs and derivations

In this Appendix, we begin by proving Proposition 2.1 (D.1). We then provide an alternative statement of the Proposition by considering sequences of $\beta_{2,n} = h_2/\sqrt{n}$, $h_2 \neq 0$, and τ_0 held fixed, and relax assumptions on the normality of the regression errors, and prove it using the same approach (Appendix D.2). Appendix D.3 laboriously derives the Hinkley distribution that characterises the sampling distribution of $\hat{\tau}$ outside of the limiting cases. We rely on the closed form to generate figures like Figure 2, or Appendix Figure A2. In Appendix D.4, we prove Corollary 3.1.

D.1 Proof of Proposition 2.1

To prove the proposition, we begin by 'recentering' the regressors, writing the regression in terms of,

$$y_i = \alpha_0 + \alpha_1[x_i - \mu] + \alpha_2[(x_i - \mu)^2 - \sigma_x^2)] + u_i$$

Now, $\alpha_1 = \beta_1 + 2\mu\beta_2$ and $\alpha_2 = \beta_2$.

The 'recentered' regressors are orthogonal, as $\mathbb{E}\left[(x_i - \mu)((x_i - \mu)^2 - \sigma_x^2)\right] = 0$ by Isserlis's theorem: odd moments of centered normal random variables are zero. We can re-parameterize the turning point now as,

$$\hat{\tau} = \mu - \frac{\hat{\alpha}_1}{2\hat{\alpha}_2}.$$

Constructing the design matrix for this regression where, $X = [1, x_i - \mu, (x_i - \mu)^2 - \sigma_x^2]$, we have that,

$$\mathbb{E}[XX'] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma_x^2 & 0 \\ 0 & 0 & 2\sigma_x^4 \end{pmatrix}$$

The asymptotic variance of the coefficients is given by $\sigma_u^2 \mathbb{E}[XX']^{-1}$, where the inverse design

matrix is simple as it is diagonal. Then, as $\beta_1 = -2\beta_2\tau_0$, we have,

$$\hat{\alpha}_1 = -2\beta_2\tau_0 + 2\mu\beta_2 + \frac{1}{\sqrt{n}}\frac{\sigma_u}{\sigma_x}Z_1, \quad \hat{\alpha}_2 = \beta_2 + \frac{1}{\sqrt{n}}\frac{\sigma_u}{\sqrt{2}\sigma_x^2}Z_2, \quad Z_1, Z_2 \sim^{\text{iid}} \mathcal{N}(0,1).$$

Simplifying,

$$\frac{\hat{\alpha}_1}{\hat{\alpha}_2} = \frac{2\beta_2(\mu - \tau_0) + \frac{1}{\sqrt{n}}\frac{\sigma_u}{\sigma_x}Z_1}{\beta_2 + \frac{1}{\sqrt{n}}\frac{\sigma_u}{\sqrt{2}\sigma_x^2}Z_2}$$

Multiplying the numerator and denominator by $\sqrt{2n}\sigma_x^2/\sigma_u$,

$$\begin{aligned}
\frac{1}{2}\frac{\hat{\alpha}_1}{\hat{\alpha}_2} &= \frac{1}{2}\frac{2\beta_2\sqrt{2n}\frac{\sigma_x^2}{\sigma_u}(\mu - \tau_0) + \sqrt{2}\sigma_x Z_1}{\beta_2\sqrt{2n}\frac{\sigma_x^2}{\sigma_u} + Z_2} \\
&= \frac{\sigma_x\sqrt{2}}{2}\frac{2\beta_2\sqrt{2n}\frac{\sigma_x^2}{\sigma_u}\left(\frac{\mu-\tau_0}{\sqrt{2}\sigma_x}\right) + Z_1}{\beta_2\sqrt{2n}\frac{\sigma_x^2}{\sigma_u} + Z_2} \\
&= \frac{\sigma_x}{\sqrt{2}}\frac{2\beta_2\sqrt{n}\frac{\sigma_x^2}{\sigma_u}\left(\frac{\mu-\tau_0}{\sigma_x}\right) + Z_1}{\beta_2\sqrt{2n}\frac{\sigma_x^2}{\sigma_u} + Z_2}
\end{aligned}$$

Defining $\theta = 2\beta_2\frac{\sigma_x}{\sqrt{2}}\sqrt{n}\frac{\sigma_x}{\sigma_u}$, we have,

$$\hat{\tau} = \mu - \frac{1}{2}\frac{\hat{\alpha}_1}{\hat{\alpha}_2} = \mu - \frac{\sigma_x}{\sqrt{2}}\frac{Z_1 + \sqrt{2}\left(\frac{\mu-\tau_0}{\sigma_x}\right)\theta}{Z_2 + \theta}, \quad Z_1, Z_2 \sim^{\text{iid}} \mathcal{N}(0,1).$$

Under the null, $\theta = 0$, so,

$$\hat{\tau}\big|_{\theta=0} \sim \mu - \frac{\sigma_x}{\sqrt{2}}\frac{Z_1}{Z_2}, \quad Z_1, Z_2 \sim^{\text{iid}} \mathcal{N}(0,1).$$

i.e.,

$$\hat{\tau} \sim \text{Cauchy}\left(\mu, \frac{\sigma_x}{\sqrt{2}}\right).$$

Rewriting $\hat{\tau}$, we can see that as $\sqrt{n} \to \infty, \theta \to \infty$,

$$\begin{aligned}
\lim_{\sqrt{n}\to\infty} \hat{\tau} &= \lim_{\theta\to\infty} \mu - \frac{\sigma_x}{\sqrt{2}} \frac{Z_1/\theta + \sqrt{2}\left(\frac{\mu-\tau_0}{\sigma_x}\right)}{Z_2/\theta + 1} \\
&= \mu - \frac{\sigma_x}{\sqrt{2}}\sqrt{2}\left(\frac{\mu-\tau_0}{\sigma_x}\right) \\
&= \tau_0
\end{aligned}$$

So $\hat{\tau}$ is root-n consistent for τ_0. Consistency allows us to take distributional limits: provided $\beta_2 \neq 0$ and τ_0 exists, the statistic is continuous in the parameters in a neighborhood of τ_0 and the continuous mapping theorem applies. To get the asymptotic variance (i.e., delta method variance), we note,

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \to^d \mathcal{N}\left(0,\ \frac{\sigma_u^2}{\sigma_x^2}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2}\sigma_x^{-2} \end{pmatrix}\right).$$

Let $g(\alpha_1, \alpha_2) = \mu - \dfrac{\alpha_1}{2\alpha_2}$. Its gradient evaluated at the true (α_1, α_2) is

$$\nabla g = \begin{pmatrix} -\dfrac{1}{2\beta_2} \\ \dfrac{\mu-\tau_0}{\beta_2} \end{pmatrix}.$$

So the delta method variance is,

$$\begin{aligned}
\begin{pmatrix} -\dfrac{1}{2\beta_2} \\ \dfrac{\mu-\tau_0}{\beta_2} \end{pmatrix}' \frac{\sigma_u^2}{\sigma_x^2}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2}\sigma_x^{-2} \end{pmatrix} \begin{pmatrix} -\dfrac{1}{2\beta_2} \\ \dfrac{\mu-\tau_0}{\beta_2} \end{pmatrix} &= \frac{\sigma_u^2}{\beta_2^2}\left(\frac{1}{4\sigma_x^2} + \frac{(\mu-\tau_0)^2}{2\sigma_x^4}\right) \\
&= \frac{1}{4}\frac{\sigma_u^2}{\beta_2^2\sigma_x^2}\left(1 + \left(\sqrt{2}\frac{\mu-\tau_0}{\sigma_x}\right)^2\right)
\end{aligned}$$

So,

$$\sqrt{n}(\hat{\tau} - \tau_0) \to^d \mathcal{N}\left(0, \frac{1}{4}\frac{\sigma_u^2}{\beta_2^2\sigma_x^2}\left(1 + \left(\sqrt{2}\frac{\mu-\tau_0}{\sigma_x}\right)^2\right)\right)$$

D.2 Asymptotic distribution under local-to-zero curvature

Proposition 2.1 in the main text derives the Hinkley distribution under fixed parameters and normal errors. We now provide a complementary asymptotic formulation using local-to-zero sequences for β_2, which delivers the same distribution as an *exact* limiting result under weaker assumptions on the error distribution (finite fourth moments rather than normality).

Proposition D.1 (Asymptotic distribution under local-to-zero curvature)**.** *Let* $\{(x_i, u_i)\}_{i=1}^n$ *be i.i.d. with*

$$x_i \sim \mathcal{N}(\mu, \sigma_x^2), \quad \mathbb{E}[u_i \mid x_i] = 0, \quad \mathbb{E}[u_i^2 \mid x_i] = \sigma_u^2, \quad \mathbb{E}[u_i^4] < \infty,$$

and suppose $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$*. Define* $\hat{\tau} = -\hat{\beta}_1/(2\hat{\beta}_2)$ *and* $\tau_0 = -\beta_1/(2\beta_2)$ *for* $\beta_2 \neq 0$*.*

1. ***Local-to-zero asymptotics (Hinkley distribution).*** *Under* $\beta_{2,n} = h_2/\sqrt{n}$ *for fixed* $h_2 \neq 0$*, with* τ_0 *held fixed (so* $\beta_{1,n} = -2\tau_0\beta_{2,n}$*):*

$$\hat{\tau} \Rightarrow \mu - \frac{\sigma_x}{\sqrt{2}} \cdot \frac{Z_1 + \left(\sqrt{2}\left(\frac{\mu-\tau_0}{\sigma_x}\right)\theta\right)}{Z_2 + \theta}, \quad Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1),$$

   where

$$\theta = h_2\sigma_x^2\sqrt{2}\,/\,\sigma_u$$

   indexes curvature signal-to-noise.

2. ***Limiting behavior indexed by*** $|\theta|$**.**

   (a) Large $|\theta|$*: The Hinkley limit concentrates at* τ_0 *with variance* $\approx \sigma_x^2/(2\theta^2)$*; approximately normal.*

   (b) Small $|\theta|$ *(*$(\mu - \tau_0)/\sigma_x$ *bounded): Approaches Cauchy*$(\mu, \sigma_x/\sqrt{2})$*, symmetric about* μ *(not* τ_0*). Delta-method inference fails, and the estimator is median biased.*

   Remark (Strong identification). *If* $\beta_2 \neq 0$ *is fixed (not local-to-zero), then* $\sqrt{n}(\hat{\tau} - \tau_0) \Rightarrow \mathcal{N}(0, V_\tau)$*: the standard delta-method result, valid when* $|\theta| \to \infty$ *with* n*.*

3. ***Observable diagnostic.*** $t_{\beta_2} \Rightarrow Z_2 + \theta \sim \mathcal{N}(\theta, 1)$*, so* $|t_{\beta_2}|$ *estimates* $|\theta|$*.*

Proof. The reparameterization used in the proof for Proposition 2.1 holds, though we no longer

have exactly normal coefficients,

$$\sqrt{n}\begin{pmatrix}\hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2\end{pmatrix} \to^d \mathcal{N}\left(0, \sigma_u^2 \begin{pmatrix}\sigma_x^2 & 0 \\ 0 & 2\sigma_x^4\end{pmatrix}\right)$$

and,

$$\hat{\alpha}_1 = \alpha_1 + \frac{1}{\sqrt{n}}\frac{\sigma_u}{\sigma_x}Z_1 + o_p(n^{-1/2})$$
$$\hat{\alpha}_2 = \alpha_2 + \frac{1}{\sqrt{n}}\frac{\sigma_u}{\sqrt{2}\sigma_x^2}Z_2 + o_p(n^{-1/2})$$

with $Z_1, Z_2 \sim^{\text{iid}} \mathcal{N}(0,1)$.

Now, with $\beta_{2,n} = h_2/\sqrt{n}$ and τ_0 fixed, the local coefficients are $\alpha_{1,n} = 2(\mu - \tau_0)h_2/\sqrt{n} \equiv h_1/\sqrt{n}$, and $\alpha_{2,n} = h_2/\sqrt{n}$. Therefore,

$$\sqrt{n}\hat{\alpha}_1 \to h_1 + \frac{\sigma_u}{\sigma_x}Z_1, \qquad \sqrt{n}\hat{\alpha}_2 \to h_2 + \frac{\sigma_u}{\sqrt{2}\sigma_x^2}Z_2.$$

We now define $\theta = h_2\sigma_x^2\sqrt{2}/\sigma_u$. Then the numerator signal-to-noise ratio is $h_1\sigma_x/\sigma_u = \sqrt{2}\,(\mu - \tau_0)/\sigma_x \cdot \theta$ (using $h_1 = 2(\mu - \tau_0)h_2$). So now we have,

$$\frac{\hat{a}_1}{\hat{a}_2} = \frac{\sqrt{n}\hat{a}_1}{\sqrt{n}\hat{a}_2} \to \sqrt{2}\sigma_x \frac{Z_1 + \sqrt{2}\left(\frac{\mu - \tau_0}{\sigma_x}\right)\theta}{Z_2 + \theta},$$

as in Proposition 2.1. Then, as we still have, $\hat{\tau} = \mu - \frac{1}{2}\frac{\hat{\alpha}_1}{\hat{\alpha}_2}$, plugging in we obtain,

$$\hat{\tau} \to \mu - \frac{\sigma_x}{\sqrt{2}} \cdot \frac{Z_1 + \sqrt{2}\left(\frac{\mu - \tau_0}{\sigma_x}\right)\theta}{Z_2 + \theta}.$$

This formulation is identical to that in Proposition 2.1, so we have the same limits: when $|\theta| \to \infty$:

$Z_2 + \theta \approx \theta$, giving $\hat{\tau} \approx \tau_0 - (\sigma_x/\sqrt{2}\theta)Z_1 \sim \mathcal{N}(\tau_0, \sigma_x^2/(2\theta^2))$. Taking $\theta \to 0$ (with $(\mu - \tau_0)/\sigma_x$ bounded), $\hat{\tau} \Rightarrow \mu - (\sigma_x/\sqrt{2})(Z_1/Z_2) \sim \text{Cauchy}(\mu, \sigma_x/\sqrt{2})$, using the fact that the ratio of two independent standard normals is Cauchy$(0, 1)$.

The t-statistic remains a useful diagnostic $t_{\beta_2} = \hat{\beta}_2/\widehat{\text{se}}(\hat{\beta}_2) = (\sqrt{2n}\sigma_x^2/\sigma_u)\hat{\beta}_2 + o_p(1)$, where the $o_p(1)$ term reflects that $\widehat{\text{se}} \xrightarrow{p} \text{se}$. Under $\beta_{2,n} = h_2/\sqrt{n}$: $t_{\beta_2} \Rightarrow Z_2 + \theta \sim \mathcal{N}(\theta, 1)$. □

D.3 Derivation of the analytic Hinkley distribution

To fully analytically characterise the Hinkley distribution, we work with the standard regressors. Since $x_i \sim \mathcal{N}(\mu, \sigma_x^2)$, when $X = [1,\ x_i,\ x_i^2]$, the design matrix

$$\mathbb{E}[XX'] = \begin{pmatrix} 1 & \mu & \mu^2 + \sigma_x^2 \\ \mu & \mu^2 + \sigma_x^2 & \mu^3 + 3\mu\sigma_x^2 \\ \mu^2 + \sigma_x^2 & \mu^3 + 3\mu\sigma_x^2 & \mu^4 + 6\mu^2\sigma_x^2 + 3\sigma_x^4 \end{pmatrix}.$$

As the errors are homoscedastic,

$$V(\beta) = \sigma_u^2 \mathbb{E}\left[X'X\right]^{-1} = \frac{\sigma_u^2}{n}\frac{1}{2\sigma_x^4}\begin{pmatrix} \mu^4 + 3\sigma_x^4 & -2\mu^3 & \mu^2 - \sigma_x^2 \\ -2\mu^3 & 2(2\mu^2 + \sigma_x^2) & -2\mu \\ \mu^2 - \sigma_x^2 & -2\mu & 1 \end{pmatrix}.$$

First, we verify Proposition 2.1: We know the ratio of two centered normals is Cauchy. What remains is to derive the parameters of the sampling distribution. The Cauchy distribution has location parameter, $-\frac{1}{2}\rho_\beta\sigma_{\beta_1}/\sigma_{\beta_2}$,

$$\rho = \frac{[V_\beta]_{23}}{\sqrt{[V_\beta]_{22}} \cdot \sqrt{[V_\beta]_{33}}}, \quad \frac{\sigma_{\beta_1}}{\sigma_{\beta_2}} = \frac{\sqrt{[V_\beta]_{22}}}{\sqrt{[V_\beta]_{33}}}$$

so, the location is,

$$-\frac{1}{2}\frac{[V_\beta]_{23}}{[V_\beta]_{33}} = -\frac{1}{2}\frac{-2\mu}{1} = \mu$$

and for the scale $\frac{1}{2}\sigma_{\beta_1}/\sigma_{\beta_2}\sqrt{1-\rho^2}$,

$$\begin{aligned}\sqrt{1-\rho^2} &= \sqrt{1 - \frac{([V_\beta]_{23})^2}{[V_\beta]_{22}[V_\beta]_{33}}} \\ &= \sqrt{1 - \frac{4\mu^2}{4\mu^2 + 2\sigma_x^2}} \\ &= \sqrt{\frac{2\sigma_x^2}{4\mu^2 + 2\sigma_x^2}}\end{aligned}$$

So the scale is,

$$\frac{1}{2}\sqrt{4\mu^2 + 2\sigma_x^2}\sqrt{\frac{2\sigma_x^2}{4\mu^2 + 2\sigma_x^2}} = \frac{\sqrt{2}}{2}\sigma_x$$

As $\sqrt{2}/2 = 1/\sqrt{2}$

$$-\frac{1}{2}\frac{\hat{\beta}_1}{\hat{\beta}_2} \sim \text{Cauchy}\left(\mu, \frac{\sigma_x}{\sqrt{2}}\right)$$

Now, when the coefficients are not mean zero, the distribution is neither Cauchy nor Gaussian. As made explicit in Hinkley (1969), the density for the ratio of two normals $r = \beta_1/\beta_2$ is,

$$\begin{aligned}f(r) = &\frac{b(r)d(r)}{\sqrt{2\pi}\sigma_{\beta_1}\sigma_{\beta_2}a(r)^3}\left[2\Phi\left(\frac{b(r)}{\sqrt{1-\rho^2}a(r)}\right) - 1\right] \\ &+ \frac{\sqrt{1-\rho^2}}{\pi\sigma_{\beta_1}\sigma_{\beta_2}a(r)^2}\exp\left(-\frac{c}{2(1-\rho^2)}\right)\end{aligned}$$

where,

$$
\begin{aligned}
a(r) &= \left(\frac{r^2}{\sigma_{\beta_1}^2} - \frac{2\rho r}{\sigma_{\beta_1}\sigma_{\beta_2}} + \frac{1}{\sigma_{\beta_2}^2} \right)^{\frac{1}{2}} \\
b(r) &= \frac{\beta_1 r}{\sigma_{\beta_1}^2} - \frac{\rho(\beta_1 + \beta_2 r)}{\sigma_{\beta_1}\sigma_{\beta_2}} + \frac{\beta_2}{\sigma_{\beta_2}^2} \\
c &= \frac{\beta_1^2}{\sigma_{\beta_1}^2} - \frac{2\rho\beta_1\beta_2}{\sigma_{\beta_1}\sigma_{\beta_2}} + \frac{\beta_2^2}{\sigma_{\beta_2}^2} \\
d(r) &= \exp\left(\frac{b(r)^2 - ca(r)^2}{2(1-\rho^2)a(r)^2} \right)
\end{aligned}
$$

In our case, we are interested in the density for $\tau = -\frac{1}{2}r$. The change of variables rule says, $f_{\hat{\tau}}(x) = f_r(g^{-1}(x)) \cdot \left| \frac{d}{dx} g^{-1}(x) \right|$ where here $g(x) = -x/2$ so, $f_\tau(x) = f(-2r)2$.

We begin by noting some useful formulas:

$$
\begin{aligned}
\sigma_{\beta_1}\sigma_{\beta_2} &= \frac{n\sigma_u^2}{2\sigma_x^4}\sqrt{2(2\mu^2 + \sigma_x^2)} \\
\frac{\rho}{\sigma_{\beta_1}\sigma_{\beta_2}} &= \frac{-2\mu}{2\mu^2 + \sigma_x^2}\frac{\sigma_x^4}{n\sigma_u^2} \\
1 - \rho^2 &= \frac{\sigma_x^2}{2\mu^2 + \sigma_x^2}.
\end{aligned}
$$

Deriving $a(x)^2$ Using our change of variables, and relabelling $n\sigma_u^2$ to just σ_u^2,

$$\begin{aligned}
a(x)^2 &= \frac{4x^2}{\sigma_{\beta_1}^2} + \frac{4\rho x}{\sigma_{\beta_1}\sigma_{\beta_2}} + \frac{1}{\sigma_{\beta_2}^2} \\
&= \frac{4x^2}{\sigma_{\beta_1}^2} + 4x\left(\frac{-2\mu\sigma_x^2}{2\mu^2+\sigma_x^2}\frac{\sigma_x^2}{\sigma_u^2}\right) + \frac{1}{\sigma_{\beta_2}^2} \\
&= \frac{4x^2}{\sigma_{\beta_1}^2} + 4x\left(\frac{-2\mu\sigma_x^2}{2\mu^2+\sigma_x^2}\frac{\sigma_x^2}{\sigma_u^2}\right) + 2\sigma_x^2\frac{\sigma_x^2}{\sigma_u^2} \\
&= \frac{4x^2}{2(2\mu^2+\sigma_x^2)}\frac{2\sigma_x^4}{\sigma_u^2} + 4x\left(\frac{-2\mu\sigma_x^2}{2\mu^2+\sigma_x^2}\frac{\sigma_x^2}{\sigma_u^2}\right) + 2\sigma_x^2\frac{\sigma_x^2}{\sigma_u^2} \\
&= \frac{2x^2}{(2\mu^2+\sigma_x^2)}\frac{2\sigma_x^4}{\sigma_u^2} + 4x\left(\frac{-2\mu\sigma_x^2}{2\mu^2+\sigma_x^2}\frac{\sigma_x^2}{\sigma_u^2}\right) + 2\sigma_x^2\frac{\sigma_x^2}{\sigma_u^2} \\
&= \frac{\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(4x^2 - 8\mu x + 2(2\mu^2+\sigma_x^2)\right) \\
&= \frac{2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2x^2 - 4\mu x + 2\mu^2 + \sigma_x^2\right)
\end{aligned}$$

We now complete the square: noting that $2x^2-4\mu x+2\mu^2 = 2(x-\mu)^2$, so that, undoing the relabel,

$$a(x)^2 = \frac{2\sigma_x^4}{n\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2(x-\mu)^2 + \sigma_x^2\right)$$

Deriving $b(x)$ We now derive $b(x)$. Here we use $\beta_1 = -2\tau\beta_2$, and again ignore n,

$$\begin{aligned}
b(x) &= -2\frac{\beta_1}{\sigma^2_{\beta_1}}x - \frac{\rho(\beta_1 - 2\beta_2 x)}{\sigma_{\beta_1}\sigma_{\beta_2}} + \frac{\beta_2}{\sigma^2_{\beta_2}} \\
&= -2\frac{-2\tau\beta_2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}x - (-2\tau\beta_2 - 2\beta_2 x)\frac{-2\mu\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)} + \frac{\beta_2}{\sigma^2_{\beta_2}} \\
&= \frac{4\tau\beta_2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}x - (\tau + x)\beta_2\frac{4\mu\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)} + \frac{\beta_2}{\sigma^2_{\beta_2}} \\
&= \frac{4\tau\beta_2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}x - (\tau + x)\beta_2\frac{4\mu\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)} + 2\beta_2\frac{\sigma_x^4}{\sigma_u^2} \\
&= \beta_2\frac{\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(4\tau x - (\tau + x)4\mu + 2(2\mu^2+\sigma_x^2)\right) \\
&= 2\beta_2\frac{\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2\tau x - (\tau + x)2\mu + 2\mu^2 + \sigma_x^2\right)
\end{aligned}$$

Noting, $2\tau x - (\tau + x)2\mu + 2\mu^2 = 2(\tau - \mu)(x - \mu)$, we have,

$$b(x) = 2\beta_2\frac{\sigma_x^4}{n\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2(\tau - \mu)(x - \mu) + \sigma_x^2\right)$$

Deriving c To obtain the constant c in the formula, ignore n again and,

$$\begin{aligned}
c &= \frac{\beta_1^2}{\sigma^2_{\beta_1}} - \frac{2\rho\beta_1\beta_2}{\sigma_{\beta_1}\sigma_{\beta_2}} + \frac{\beta_2^2}{\sigma^2_{\beta_2}} \\
&= \frac{4\tau^2\beta_2^2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)} - -4\tau\beta_2^2\frac{-2\mu\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)} + \frac{\beta_2^2}{\sigma^2_{\beta_2}} \\
&= \frac{4\tau^2\beta_2^2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)} - 8\tau\beta_2^2\frac{\mu\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)} + \beta_2^2 2\frac{\sigma_x^4}{\sigma_u^2} \\
&= 2\beta_2^2\frac{\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2\tau^2 - 4\tau\mu + 2\mu^2 + \sigma_x^2\right)
\end{aligned}$$

As $2\tau^2 - 4\tau\mu + 2\mu^2 = 2(\tau - \mu)^2$, we have,

$$c = 2\beta_2^2 \frac{\sigma_x^4}{n\sigma_u^2(2\mu^2 + \sigma_x^2)} \left(2(\tau - \mu)^2 + \sigma_x^2\right)$$

Deriving $d(x)$ So, first define,

$$K = \frac{\sigma_x^4}{\sigma_u^2(2\mu^2 + \sigma_x^2)}$$

and again subsuming n in σ_u^2 for the algebra,

$$\begin{aligned} d(x) &= \exp\left(\frac{b(x)^2 - ca(x)^2}{2(1-\rho^2)a(x)^2}\right) \\ &= \exp\left(\frac{\left[2\beta_2 K\left(2(\tau-\mu)(x-\mu) + \sigma_x^2\right)\right]^2 - \left[2\beta_2^2 K\left(2(\tau-\mu)^2 + \sigma_x^2\right)\right]\left[2K\left(2(x-\mu)^2 + \sigma_x^2\right)\right]}{2(1-\rho^2)\left[2K\left(2(x-\mu)^2 + \sigma_x^2\right)\right]}\right) \\ &= \exp\left(2\beta_2^2 K \frac{\left[\left(2(\tau-\mu)(x-\mu) + \sigma_x^2\right)\right]^2 - \left[\left(2(\tau-\mu)^2 + \sigma_x^2\right)\right]\left[\left(2(x-\mu)^2 + \sigma_x^2\right)\right]}{2(1-\rho^2)\left[2\left(2(x-\mu)^2 + \sigma_x^2\right)\right]}\right) \end{aligned}$$

The numerator simplifies,

$$\left[\left(2(\tau-\mu)(x-\mu) + \sigma_x^2\right)\right]^2 - \left[\left(2(\tau-\mu)^2 + \sigma_x^2\right)\right]\left[\left(2(x-\mu)^2 + \sigma_x^2\right)\right] = -2\sigma_x^2(\tau - x)^2$$

And note that $1 - \rho^2 = \frac{\sigma_x^2}{2\mu^2 + \sigma_x^2} = K\sigma_u^2/\sigma_x^2$ so we have,

$$d(x) = \exp\left(\beta_2^2 K \frac{-2\sigma_x^2(\tau - x)^2}{K\sigma_u^2/\sigma_x^2\left[2K\left(2(x-\mu)^2 + \sigma_x^2\right)\right]}\right)$$

which gives, undoing the σ_u^2 relabel,

$$d(x) = \exp\left(-2\beta_2^2 \frac{\sigma_x^4}{n\sigma_u^2} \frac{(\tau - x)^2}{\left[(2(x-\mu)^2 + \sigma_x^2)\right]}\right)$$

Taking stock, we have,

$$
\begin{aligned}
a(x)^2 &= \frac{2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2(x-\mu)^2+\sigma_x^2\right) \\
b(x) &= 2\beta_2\frac{\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2(\tau-\mu)(x-\mu)+\sigma_x^2\right) \\
c &= 2\beta_2^2\frac{\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2(\tau-\mu)^2+\sigma_x^2\right) \\
d(x) &= \exp\left(-2\beta_2^2\frac{\sigma_x^4}{\sigma_u^2}\frac{(\tau-x)^2}{[(2(x-\mu)^2+\sigma_x^2)]}\right)
\end{aligned}
$$

Now, noting that $\theta^2 = 2\beta_2^2\frac{n\sigma_x^4}{\sigma_u^2}$, we can write these as,

$$
\begin{aligned}
a(x)^2 &= \frac{2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2(x-\mu)^2+\sigma_x^2\right) \\
b(x) &= \theta^2\frac{1}{\beta_2(2\mu^2+\sigma_x^2)}\left(2(\tau-\mu)(x-\mu)+\sigma_x^2\right) \\
c &= \theta^2\frac{1}{(2\mu^2+\sigma_x^2)}\left(2(\tau-\mu)^2+\sigma_x^2\right) \\
d(x) &= \exp\left(-\theta^2\frac{(\tau-x)^2}{[(2(x-\mu)^2+\sigma_x^2)]}\right)
\end{aligned}
$$

Returning to the Hinkley distribution, we can simplify the ρ terms, and the σ terms to obtain,

$$
\begin{aligned}
f_\tau(x) = &\frac{2b(x)d(x)}{\sqrt{2\pi}\left(n\frac{\sigma_u^2}{2\sigma_x^4}\sqrt{2(2\mu^2+\sigma_x^2)}\right)a(x)^3}\left[2\Phi\left(\frac{b(x)}{\sqrt{1-\rho^2}a(x)}\right)-1\right] \\
&+2\frac{\sqrt{\frac{\sigma_x^2}{2\mu^2+\sigma_x^2}}}{\pi\left(n\frac{\sigma_u^2}{2\sigma_x^2}\sqrt{2(2\mu^2+\sigma_x^2)}\right)a(x)^2}\exp\left(-\frac{c}{2}\left(1+2\left(\frac{\mu}{\sigma_x}\right)^2\right)\right)
\end{aligned}
$$

Cleaning up,

$$f_\tau(x) = \underbrace{\frac{2b(x)d(x)}{\sqrt{2\pi}\left(\frac{\sigma_u^2}{2\sigma_x^2}\sqrt{2(2\mu^2+\sigma_x^2)}\right)a(x)^3}\left[2\Phi\left(\frac{b(x)}{\sqrt{1-\rho^2}a(x)}\right)-1\right]}_{\text{term 1}} + \underbrace{\frac{\sigma_x/\sqrt{2}}{\pi\left(\frac{\sigma_u^2}{2\sigma_x^4}(2\mu^2+\sigma_x^2)\right)a(x)^2}\exp\left(-\frac{c}{2}\left(1+2\left(\frac{\mu}{\sigma_x}\right)^2\right)\right)}_{\text{term 2}}$$

Plugging in for c and $a(x)^2$ in the second term,

$$\text{term 2} = 2\frac{\sigma_x/\sqrt{2}}{\pi(2(x-\mu)^2+\sigma_x^2)}\exp\left(-\frac{1}{2}\theta^2\frac{1}{(2\mu^2+\sigma_x^2)}\left(2(\tau-\mu)^2+\sigma_x^2\right)\left(1+2\left(\frac{\mu}{\sigma_x}\right)^2\right)\right)$$

Cleaning up the second term,

$$\text{term 2} = \frac{\sigma_x/\sqrt{2}}{\pi((x-\mu)^2+\sigma_x^2/2)}\exp\left(-\theta^2\frac{1}{(\mu^2+\sigma_x^2/2)}\left(2(\tau-\mu)^2+\sigma_x^2\right)\left(1+2\left(\frac{\mu}{\sigma_x}\right)^2\right)\right)$$

Simplifying further,

$$\begin{aligned}\text{term 2} &= \frac{\sigma_x/\sqrt{2}}{\pi((x-\mu)^2+\sigma_x^2/2)}\exp\left(-\theta^2\frac{2}{2\mu^2+\sigma_x^2}\left(2(\tau-\mu)^2+\sigma_x^2\right)\left(\frac{\sigma_x^2+2\mu^2}{\sigma_x^2}\right)\right)\\ &= \frac{\sigma_x/\sqrt{2}}{\pi((x-\mu)^2+\sigma_x^2/2)}\exp\left(-\frac{\theta^2}{2}\left(1+\left(\frac{2(\tau-\mu)}{\sigma_x}\right)^2\right)\right)\end{aligned}$$

Define,

$$\kappa^2 = \theta^2\left(1+\left(\sqrt{2}\frac{\tau-\mu}{\sigma_x}\right)^2\right)$$

So,

$$\text{term 2} = \frac{\sigma_x/\sqrt{2}}{\pi((x-\mu)^2+\sigma_x^2/2)}\exp\left(-\kappa^2/2\right)$$

Now reexamine term 1,

$$\text{term 1} = \frac{2b(x)}{\sqrt{2\pi}\left(\frac{\sigma_u^2}{2\sigma_x^4}\sqrt{2(2\mu^2+\sigma_x^2)}\right)a(x)^3}\left[2\Phi\left(\frac{b(x)}{\sqrt{1-\rho^2}a(x)}\right)-1\right]d(x)$$

Consider $b(x)/a(x)^2$. In our context, the sign of $b(x)$ is undone by the odd function $2\Phi() - 1$ so we can put absolute values over any signed quantities,

$$\begin{aligned}
b(x)/a(x)^2 &= \frac{\theta^2\frac{1}{|\beta_2|(2\mu^2+\sigma_x^2)}\left(2(\tau-\mu)(x-\mu)+\sigma_x^2\right)}{\frac{2\sigma_x^4}{\sigma_u^2(2\mu^2+\sigma_x^2)}\left(2(x-\mu)^2+\sigma_x^2\right)} \\
&= \frac{\theta^2\frac{1}{(2\mu^2+\sigma_x^2)}\left(2(\tau-\mu)(x-\mu)+\sigma_x^2\right)}{\sqrt{2}|\beta_2|\frac{\sigma_x^2}{\sigma_u}\frac{\sqrt{2}\sigma_x^2}{\sigma_u(2\mu^2+\sigma_x^2)}\left(2(x-\mu)^2+\sigma_x^2\right)} \\
&= \frac{\theta^2\left(2(\tau-\mu)(x-\mu)+\sigma_x^2\right)}{|\theta|\frac{\sqrt{2}\sigma_x^2}{\sigma_u}\left(2(x-\mu)^2+\sigma_x^2\right)} \\
&= |\theta|\frac{\sigma_u}{\sqrt{2}\sigma_x^2}\frac{\left(2(\tau-\mu)(x-\mu)+\sigma_x^2\right)}{\left(2(x-\mu)^2+\sigma_x^2\right)}
\end{aligned}$$

Now note that,

$$a(x) = \frac{\sqrt{2}\sigma_x^2}{\sigma_u\sqrt{(2\mu^2+\sigma_x^2)}}\sqrt{(2(x-\mu)^2+\sigma_x^2)}$$

So,

$$\begin{aligned}
b(x)/a(x)^3 &= |\theta|\frac{\sigma_u}{\sqrt{2}\sigma_x^2}\frac{\left(2(\tau-\mu)(x-\mu)+\sigma_x^2\right)}{\left(2(x-\mu)^2+\sigma_x^2\right)}\frac{1}{\frac{\sqrt{2}\sigma_x^2}{\sigma_u\sqrt{(2\mu^2+\sigma_x^2)}}\sqrt{(2(x-\mu)^2+\sigma_x^2)}} \\
&= |\theta|\frac{1}{2}\frac{\sigma_u^2}{\sigma_x^4}\frac{\left(2(\tau-\mu)(x-\mu)+\sigma_x^2\right)}{\left(2(x-\mu)^2+\sigma_x^2\right)}\frac{\sqrt{(2\mu^2+\sigma_x^2)}}{\sqrt{(2(x-\mu)^2+\sigma_x^2)}}
\end{aligned}$$

So then,

$$\begin{aligned}
&\frac{2b(x)}{\sqrt{2\pi}\left(\frac{\sigma_u^2}{2\sigma_x^4}\sqrt{2(2\mu^2+\sigma_x^2)}\right)a(x)^3} \\
&= \frac{2\sigma_x^4}{\sqrt{\pi}\sigma_u^2\sqrt{2\mu^2+\sigma_x^2}}\frac{b(x)}{a(x)^3} \\
&= \frac{2\sigma_x^4}{\sqrt{\pi}\sigma_u^2\sqrt{2\mu^2+\sigma_x^2}}|\theta|\frac{1}{2}\frac{\sigma_u^2}{\sigma_x^4}\frac{(2(\tau-\mu)(x-\mu)+\sigma_x^2)}{(2(x-\mu)^2+\sigma_x^2)}\frac{\sqrt{(2\mu^2+\sigma_x^2)}}{\sqrt{(2(x-\mu)^2+\sigma_x^2)}} \\
&= \frac{1}{\sqrt{\pi}}|\theta|\frac{(2(\tau-\mu)(x-\mu)+\sigma_x^2)}{(2(x-\mu)^2+\sigma_x^2)^{\frac{3}{2}}}
\end{aligned}$$

Or,

$$= \frac{1}{\sqrt{\pi}}\frac{(2(\tau-\mu)(x-\mu)+\sigma_x^2)}{\frac{\sigma_x}{\kappa}(2(\tau-\mu)^2+\sigma_x^2)^{\frac{1}{2}}\,(2(x-\mu)^2+\sigma_x^2)^{\frac{3}{2}}}$$

Now $d(x)$ can be written as,

$$\begin{aligned}
d(x) &= \exp\left(-\theta^2(\frac{(\tau-x)^2}{2(x-\mu)^2+\sigma_x^2}\right) \\
&= \exp\left(-\theta^2\left(1+2\left(\frac{(\tau-\mu)}{\sigma_x}\right)^2\right)\frac{(\tau-x)^2}{2(x-\mu)^2+\sigma_x^2}\right) \\
&= \exp\left(-\frac{1}{2}\frac{(\tau-x)^2}{\frac{1}{2\kappa^2}\left(1+2\left(\frac{\tau-\mu}{\sigma_x}\right)^2\right)}\frac{1}{(2(x-\mu)^2+\sigma_x^2)}\right) \\
&= \exp\left(-\frac{1}{2}\frac{(\tau-x)^2}{\frac{\sigma_x^2}{2\kappa^2}\left(1+\left(\sqrt{2}\frac{\tau-\mu}{\sigma_x}\right)^2\right)^2}\right) \\
&= \exp\left(-\frac{1}{2}\left(\frac{(\tau-x)}{\frac{\sigma_x}{\sqrt{2}|\theta|}\sqrt{1+\left(\sqrt{2}\frac{\tau-\mu}{\sigma_x}\right)^2}}\right)^2\right)
\end{aligned}$$

So the entire term 1 becomes,

$$\text{term } 1 = \frac{1}{\sqrt{\pi}} \frac{(2(\tau-\mu)(x-\mu)+\sigma_x^2)}{\frac{\sigma_x}{\kappa}(2(\tau-\mu)^2+\sigma_x^2)^{\frac{1}{2}}\left(2(x-\mu)^2+\sigma_x^2\right)^{\frac{3}{2}}} \exp\left(-\frac{1}{2}\left(\frac{(\tau-x)}{\frac{\sigma_x}{\sqrt{2}|\theta|}\sqrt{1+\left(\sqrt{2}\frac{\tau-\mu}{\sigma_x}\right)}}\right)^2\right)$$

where I have omitted the $2\Phi() - 1$ term. This peaks at τ and is approximately normal with,

$$\kappa \to \infty \implies \hat{\tau} \sim \mathcal{N}\left(\tau, \frac{\sigma_x^2}{2}\theta^{-2}\left(1+\sqrt{2}\frac{\tau-\mu}{\sigma_x}\right)\right)$$

D.4 Proof of Corollary 3.1

We begin with recentering the regressors so the regression is as follows,

$$y_i = a_0 + \underbrace{(\beta_1 + 2\mu\beta_2)}_{\equiv\alpha_1}[x_i - \mu] + \beta_2\left[(x-\mu)^2 - \sigma_x^2\right] + \eta_1 x_i w_i + u_i$$

and now the first two regressors are orthogonal. Then, for $X_i = (1, x_i - \mu, (x-\mu)^2 - \sigma_x^2, x_i w_i)$,

$$\mathbb{E}[X_i X_i'] = \begin{pmatrix} 1 & 0 & 0 & \rho\sigma_x \\ 0 & \sigma_x^2 & 0 & \mu\rho\sigma_x \\ 0 & 0 & 2\sigma_x^4 & 2\rho\sigma_x^3 \\ \rho\sigma_x & \mu\rho\sigma_x & 2\rho\sigma_x^3 & \mu^2 + \sigma_x^2(1+2\rho^2) \end{pmatrix}$$

which relies on Isserlis's theorem for the higher moments. Consider the off-diagonal entries. First,

$$\begin{aligned} \text{Cov}(x-\mu, xw) &= \mathbb{E}[(x-\mu)(x-\mu+\mu)w] \\ &= \mathbb{E}[(x-\mu)^2 w] + \mu\mathbb{E}[(x-\mu)w] \\ &= \mu\rho\sigma_x \end{aligned}$$

where $E[(x-\mu)^2 w] = 0$ by Isserlis's theorem. For,

$$\begin{aligned}
\text{Cov}((x-\mu)^2 - \sigma_x^2, xw) &= \mathbb{E}[((x-\mu)^2 - \sigma_x^2)(x-\mu+\mu)w] \\
&= \mathbb{E}[((x-\mu)^2 - \sigma_x^2)(x-\mu)w] + \mu\mathbb{E}[((x-\mu)^2 - \sigma_x^2)w] \\
&= \mathbb{E}[(x-\mu)^3 w] - \sigma_x^2 \mathbb{E}[(x-\mu)w] \\
&= \mathbb{E}[(x-\mu)^3 w] - \sigma_x^3 \rho \\
&= 2\rho\sigma_x^3
\end{aligned}$$

as $\mathbb{E}[((x-\mu)^2 - \sigma_x^2)w] = 0$ by Isserlis's theorem; likewise, for the last line, we can compute all the pairings for even moments to conclude $\mathbb{E}[(x-\mu)^3 w] = 3\rho\sigma_x^3$. Finally,

$$\begin{aligned}
\mathbb{E}[x_i^2 w_i^2] &= \mathbb{E}[(x_i - \mu + \mu)^2 w_i^2] \\
&= \mathbb{E}[(x-\mu)^2 w^2 + 2\mu(x-\mu)w^2 + \mu^2 w^2] \\
&= \mathbb{E}[(x-\mu)^2 w^2] + \mu^2
\end{aligned}$$

as $2\mu\mathbb{E}[(x-\mu)w^2] = 0$ again by Isserlis's. Isserlis also gives,

$$\begin{aligned}
\mathbb{E}[(x-\mu)^2 w^2] &= \mathbb{E}[(x-\mu)^2]\mathbb{E}[w^2] + 2\mathbb{E}[(x-\mu)w]^2 \\
&= \sigma_x^2 + 2(\rho\sigma_x)^2 \\
&= \sigma_x^2(1 + 2\rho^2)
\end{aligned}$$

This completes our characterization of the design matrix. Using a symbolic mathematical toolkit (`sympy`, in Python), we invert the design matrix and derive the variance-covariance matrix for the OLS coefficients,

$$\sigma_u^2\mathbb{E}[X_iX_i']^{-1} = \begin{pmatrix} \frac{\sigma_u^2\left(\mu^2\rho^2-\mu^2-\sigma_x^2\right)}{\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2} & -\frac{\mu\rho^2\sigma_u^2}{\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2} & -\frac{\rho^2\sigma_u^2}{\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2} & \frac{\rho\sigma_u^2\sigma_x}{\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2} \\ -\frac{\mu\rho^2\sigma_u^2}{\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2} & \frac{\sigma_u^2\left(-\mu^2+\rho^2\sigma_x^2-\sigma_x^2\right)}{\sigma_x^2(\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2)} & -\frac{\mu\rho^2\sigma_u^2}{\sigma_x^2(\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2)} & \frac{\mu\rho\sigma_u^2}{\sigma_x(\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2)} \\ -\frac{\rho^2\sigma_u^2}{\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2} & -\frac{\mu\rho^2\sigma_u^2}{\sigma_x^2(\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2)} & \frac{\sigma_u^2\left(\mu^2\rho^2-\mu^2-\rho^2\sigma_x^2-\sigma_x^2\right)}{2\sigma_x^4(\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2)} & \frac{\rho\sigma_u^2}{\sigma_x(\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2)} \\ \frac{\rho\sigma_u^2\sigma_x}{\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2} & \frac{\mu\rho\sigma_u^2}{\sigma_x(\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2)} & \frac{\rho\sigma_u^2}{\sigma_x(\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2)} & -\frac{\sigma_u^2}{\mu^2\rho^2-\mu^2+\rho^2\sigma_x^2-\sigma_x^2} \end{pmatrix}$$

Now,

$$\hat{\tau}(w) = -\frac{1}{2}\frac{\hat{\alpha}_1 - 2\mu\hat{\beta}_2 + \hat{\eta}_1 w}{\hat{\beta}_2}$$

To construct the null, define matrix,

$$V = \begin{pmatrix} 0 & 1 & -2\mu & w \\ 0 & 0 & -2 & 0 \end{pmatrix}$$

Under the null, these are distributed,

$$\begin{pmatrix} \hat{\alpha}_1 - 2\mu\hat{\beta}_2 + \hat{\eta}_1 w \\ -2\hat{\beta}_2 \end{pmatrix} \sim \mathcal{N}\left(0, \Sigma\right), \quad \Sigma = V\sigma_u^2\mathbb{E}[X_iX_i']^{-1}V'$$

where,

$$\begin{aligned} \Sigma_{11} &= \frac{\sigma_u^2\left(2\mu^4\rho^2 - 2\mu^4 + 2\mu^2\rho^2\sigma_x^2 - 3\mu^2\sigma_x^2 - 2\mu\rho\sigma_x^3 w + \rho^2\sigma_x^4 - \sigma_x^4 w^2 - \sigma_x^4\right)}{\sigma_x^4\left(\mu^2\rho^2 - \mu^2 + \rho^2\sigma_x^2 - \sigma_x^2\right)} \\ \Sigma_{12} &= \frac{2\sigma_u^2\left(\mu^3\rho^2 - \mu^3 - \mu\sigma_x^2 - \rho\sigma_x^3 w\right)}{\sigma_x^4\left(\mu^2\rho^2 - \mu^2 + \rho^2\sigma_x^2 - \sigma_x^2\right)} \\ \Sigma_{22} &= \frac{2\sigma_u^2\left(\mu^2\rho^2 - \mu^2 - \rho^2\sigma_x^2 - \sigma_x^2\right)}{\sigma_x^4\left(\mu^2\rho^2 - \mu^2 + \rho^2\sigma_x^2 - \sigma_x^2\right)} \end{aligned}$$

Using these, and recalling, $\mathbb{E}[x \mid w] = \mu + \rho\sigma_x w$ we have,

$$\begin{aligned}
\breve{\mu}(w) &= \frac{\Sigma_{12}}{\Sigma_{22}} \\
&= \frac{\mu^3\rho^2 - \mu^3 - \mu\sigma_x^2 - \rho\sigma_x^3 w}{\mu^2\rho^2 - \mu^2 - \rho^2\sigma_x^2 - \sigma_x^2} \\
&= \frac{-\mu^3\rho^2 + \mu^3 + \mu\sigma_x^2 + \rho\sigma_x^3 w}{-\mu^2\rho^2 + \mu^2 + \rho^2\sigma_x^2 + \sigma_x^2} \\
&= \frac{\mu(\mu^2(1-\rho^2)) + \mu\sigma_x^2 + \sigma_x^2(\mathbb{E}[x \mid w] - \mu)}{\mu^2(1-\rho^2) + \sigma_x^2(1+\rho^2)} \\
&= \frac{(\mu^2(1-\rho^2))}{\mu^2(1-\rho^2) + \sigma_x^2(1+\rho^2)}\mu + \frac{\sigma_x^2\mathbb{E}[x \mid w]}{\mu^2(1-\rho^2) + \sigma_x^2(1+\rho^2)}
\end{aligned}$$

Defining,

$$\lambda = \frac{\sigma_x^2(1+\rho^2)}{\mu^2(1-\rho^2) + \sigma_x^2(1+\rho^2)}$$

Note that λ can bet written as,

$$\lambda = \frac{(1+\rho^2)}{(\mu/\sigma_x)^2(1-\rho^2) + (1+\rho^2)}$$

which is clearly governed by how strong the w signal is (ρ^2) relative to the strength of the baseline mean after conditioning on w.

Now, we have,

$$\breve{\mu}(w) = \lambda\frac{\mathbb{E}[x \mid w]}{1+\rho^2} + (1-\lambda)\mu$$

For the scale parameter, we must solve for,

$$\breve{\sigma}(w) = \frac{\Sigma_{11}}{\Sigma_{22}}\sqrt{1 - \frac{(\Sigma_{12})^2}{\Sigma_{11}\Sigma_{22}}}$$

We derive this with `sympy`,

$$\breve{\sigma}(w)^2 = \frac{\sigma_x^2}{2} \cdot \frac{\mathbb{E}[x^2](1-\rho^2)}{\mu^2(1-\rho^2) + \sigma_x^2(1+\rho^2)} \cdot \left(1 + \frac{(\sigma_x w - \mu\rho)^2}{\mu^2(1-\rho^2) + \sigma_x^2(1+\rho^2)}\right)$$

We can write this in terms of λ,

$$\breve{\sigma}(w) = \frac{\sigma_x}{\sqrt{2}} \cdot \frac{1}{1+\rho^2} \cdot \sqrt{1+\rho^2(1-2\lambda)} \cdot \sqrt{1+\rho^2+\lambda\left(w-\rho\frac{\mu}{\sigma_x}\right)^2}$$

D.5 Extension to higher-order polynomials and non-Gaussian regressors

Proposition 2.1 establishes an exact distributional result for the quadratic case ($K = 2$) with Gaussian regressors. This appendix extends the analysis to polynomial models of arbitrary order $K \geq 2$ and to non-Gaussian regressor distributions, showing that the same ratio-of-normals structure governs the turning-point estimator, but with additional distortion terms that are exactly zero in the quadratic-Gaussian baseline.

Corollary D.2 (Turning-point inference for higher-order polynomials)**.** *Under the conditions of Proposition 2.1, replace the quadratic model with a polynomial of order $K \geq 2$:*

$$y_i = \sum_{k=0}^{K} \beta_k x_i^k + u_i,$$

and let $f(x) = \sum_{k=0}^{K} \beta_k x^k$ have a turning point τ_0 satisfying $f'(\tau_0) = 0$ and $f''(\tau_0) \neq 0$. Let $\hat{\tau}$ solve $\hat{f}'(\hat{\tau}) = 0$ in a neighborhood of τ_0. Then:

$$\hat{\tau} = \mu - H_K \cdot C_K + \tilde{R}_n$$

where $Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and:

1. *H_K: **Hinkley term.***

$$H_K = s_K \cdot \frac{Z_1 + \delta_K \theta_K(\mu)}{\theta_K(\mu) + Z_2}$$

is a ratio of independent normal random variables, with $\theta_K(\mu) = f''(\mu)/\text{se}(\hat{f}''(\mu))$ the curvature t-statistic at μ; $\delta_K = (\mu - \tau_0)/s_K$ the displacement of the turning point from μ in scale units; and $s_K = \text{se}(\hat{f}'(\mu))/\text{se}(\hat{f}''(\mu))$ the ratio of standard errors. When $K = 2$: $s_K = \sigma_x/\sqrt{2}$, $\theta_K(\mu) = \theta$, and $\delta_K\theta_K(\mu) = \sqrt{2}(\mu - \tau_0)/\sigma_x \times \theta$, recovering Proposition 2.1.

2. *C_K: **Correlation correction.***

$$C_K = \frac{\omega_K + Z_2}{\omega_K + \rho_K Z_1 + \sqrt{1-\rho_K^2}\, Z_2}$$

where $\rho_K = \mathrm{Corr}(\hat{f}'(\mu), \hat{f}''(\mu))$. When $\rho_K = 0$, $C_K = 1$ identically and the Proposition 2.1 distribution obtains exactly. For $K = 2$ with Gaussian x, $\rho_K = 0$ by the Isserlis orthogonality used in the proof of Proposition 2.1 ($E[(x-\mu)^3] = 0$).

3. $\tilde{R}_n$***: Taylor remainder.***

$$\tilde{R}_n = O_p\left(\frac{f'''(\tau_0)}{f''(\tau_0)}\right)$$

under weak identification, and $O_p(n^{-1})$ under strong identification. For $K = 2$, $\tilde{R}_n = 0$ exactly since $f''' \equiv 0$.

4. ***Observable diagnostic.*** *The feasible curvature t-statistic $\hat{\theta}_K(\hat{\mu}) = \hat{f}''(\hat{\mu})/\widehat{\mathrm{se}}(\hat{f}''(\hat{\mu}))$ satisfies $\hat{\theta}_K(\hat{\mu}) = \theta_K(\mu) + O_p(1)$, and correctly classifies the identification regime: large $|\hat{\theta}_K(\mu)|$ implies strong identification (delta-method inference valid), small $|\hat{\theta}_K(\mu)|$ signals the Cauchy-like pathologies of Proposition 2.1.*

Proof sketch. The estimator $\hat{\tau}$ solves $\hat{f}'(\hat{\tau}) = 0$. Expanding around μ:

$$0 = \hat{f}'(\mu) + \hat{f}''(\mu)(\hat{\tau} - \mu) + \tfrac{1}{2}\hat{f}'''(\mu)(\hat{\tau} - \mu)^2 + \cdots$$

The first-order Newton step gives $\hat{\tau} \approx \mu - \hat{f}'(\mu)/\hat{f}''(\mu)$. Both $\hat{f}'(\mu) = \sum_k k\hat{\beta}_k \mu^{k-1}$ and $\hat{f}''(\mu) = \sum_k k(k-1)\hat{\beta}_k \mu^{k-2}$ are linear in $\hat{\beta}$, hence jointly normal under Gaussian errors. Writing their joint distribution as

$$\hat{f}'(\mu) = f'(\mu) + \sigma_1 Z_1, \qquad \hat{f}''(\mu) = f''(\mu) + \sigma_2(\rho_K Z_1 + \sqrt{1-\rho_K^2}\, Z_2)$$

where $\sigma_j = \mathrm{se}(\hat{f}^{(j)}(\mu))$, and using $f'(\tau_0) = 0$ to expand $f'(\mu) = f''(\tau_0)(\mu - \tau_0) + O(f'''(\tau_0)(\mu - \tau_0)^2)$, the ratio decomposes as $\hat{\tau} - \mu = -H_K \cdot C_K + \tilde{R}_n$.

For $K = 2$, three simplifications occur simultaneously. First, the recentered basis $\{1,\ x_i - \mu,\ (x_i - \mu)^2 - \sigma_x^2\}$ diagonalizes under Gaussian x via Isserlis's theorem ($E[(x-\mu)^3] = 0$), giving $\rho_K = 0$ and hence $C_K = 1$. Second, $f''' \equiv 0$ for a quadratic, so both the Newton remainder and the noncentrality residual vanish: $\tilde{R}_n = 0$. Third, the diagonal of $(X'X)^{-1}$ in the recentered basis yields $s_K = \sigma_x/\sqrt{2}$. Together these recover Proposition 2.1. □

Two sources of departure from Proposition 2.1. The correlation ρ_K can be nonzero through two distinct channels:

1. *Higher-order polynomials ($K > 2$, even with Gaussian x).* Cross-moments of higher-order basis functions do not vanish: for $K \geq 3$, $\hat{f}'(\mu)$ and $\hat{f}''(\mu)$ depend on overlapping linear combinations of $\hat{\beta}$, and the off-diagonal elements of $(X'X)^{-1}$ are generically nonzero. The Taylor remainder $\tilde{R}_n$ is also nonzero since $f''' \neq 0$.

2. *Non-Gaussian regressors (even with $K = 2$).* The Isserlis orthogonality requires $E[(x - \mu)^3] = 0$, i.e. zero skewness. This holds for any symmetric distribution but fails for asymmetric ones. In particular, pooling data across heterogeneous climate regimes produces a bimodal regressor distribution with $\kappa_3 = E[(x - \mu)^3] \neq 0$ whenever the mixture is asymmetric, giving $\rho_K \neq 0$ and $C_K \neq 1$ even for a quadratic model.

In climate applications, both channels typically operate simultaneously: flexible polynomial specifications are estimated on pooled data from heterogeneous regimes. The Gaussian-quadratic case of Proposition 2.1 is therefore a knife-edge — any departure breaks the independence that makes the result exact.

Implications for adaptation tests. When data are pooled across regimes, the expansion point μ in Corollary D.2 is the *global* mean of the regressor. The H_K term then pulls each regime's turning-point estimate toward $\bar{\mu}$ rather than toward the local regime mean μ_r. Estimated turning points from a pooled higher-order polynomial will therefore appear to *converge* across regimes — a pattern that could be misinterpreted as partial adaptation, but which is purely an artifact of the median-bias mechanism operating at the global scale. By contrast, separate quadratic estimates within each regime are biased toward local means, generating the spurious adaptation pattern described in Section 2.1. Neither pattern is informative about true adaptation; both are consequences of the ratio-of-normals pathology under weak curvature.

E Empirical Bayes implementation details

This appendix details the empirical Bayes procedure we implement in Section 3.2.

In contrast to the Bayesian estimates of the preceding section which simply relied on a fixed prior, *empirical* Bayes methods use the data to estimate the prior, which we refer to as $G()$. There are many interpretations of these estimators: e.g., to estimate a hierarchical model, to solve for decision rules that minimize their compound risk, to mimic an 'oracle.' See, e.g., Efron (2019) and Koenker and Gu (2024) or Walters (2024) for exposition.

In our setting, standard Empirical Bayes estimators, like a normal-likelihood, normal-prior model are inappropriate because the sampling distribution of the estimators $\hat{\tau}_i$ is considerably skew, and we have no reason to believe the empirical prior on τ ought to be normal. Instead, we proceed by following state-of-the-art methods in nonparametric maximum likelihood estimation of the prior $G()$. We model $G()$ as a discrete mixture over a fine mesh of potential peaks, τ_j

$$G(\tau) = \sum_{j=1}^{J} \phi_j \delta(\tau - \tau_j)$$

where $\delta()$ is the Dirac delta function. This is sometimes referred to as a 'spiky' prior.

We recover mixture weights $\phi_j > 0$ by solving,

$$\min_{\lambda_j} \; -\sum_i \log\left(\sum_j \lambda_j L_{ij}\right) \quad \text{subject to} \quad \lambda_j > 0, \sum_j \lambda_j = 1$$

The idea is that is substantially easier to fit high-dimensional mixtures on a discretized grid using tools from convex optimization (Koenker and Mizera, 2014) by recovering mixture weights ϕ_j than it is to fit low-dimensional mixture models by, e.g., the expectation-maximisation (EM) algorithm (Dempster et al., 1977). To solve this optimization problem, we rely on the toolkit from Youngseok Kim et al. (2020). In practice, we use a $J = 200$ point grid for τ on the 0.5th to 99.5th quantiles of

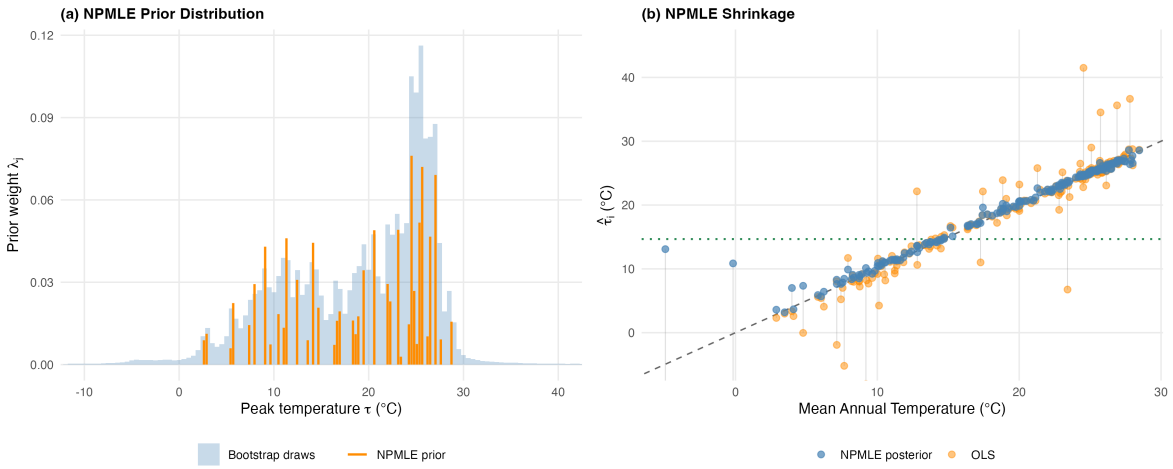


Figure A15: **Estimation of the NPMLE τ_i prior and resulting posterior means for GDP**
Panel (a) shows the parametric bootstrap distribution of τ_i draws across all $(\hat{\beta}_{1i}, \hat{\beta}_{2i})$ in blue, and the estimated prior $G()$. The right panel shows posterior means (blue) and OLS point estimates (orange). Little shrinkage occurs.

the draws, in degrees C.

The elements of L_{ij} are the likelihoods $p_i(\hat{\tau}_i - \tau_j)$, where we fit $p_i(\hat{\tau}_i - \tau)$ by fitting a mixture of BIC-selected gaussian distributions, $\hat{p}_i()$, by EM (we use `densityMclust` in `R`), onto the residuals R_i of bootstrap draws b from the likelihood for $\hat{\tau}_i$; $R_i = \tau_i^b - \hat{\tau}_i$.

Appendix Figure A15 shows the estimated prior for the temperature-GDP data against the bootstrap distribution of $\hat{\tau}_i$ (left panel). The right panel shows the posterior means of τ_i. The sampling distributions for each $\hat{\tau}_i$ are far too diffuse to discipline a prior. The resulting spiky prior estimate is diffuse across the distribution of $\hat{\tau}_i$ and mirrors its density. Subsequently, posterior means essentially align with the medians of the likelihood, simply shrinking OLS outliers towards the 45 degree line.

Appendix Figure A16 shows the same plot for mortality. The estimated prior performs equally as poorly as it did for the GDP data.

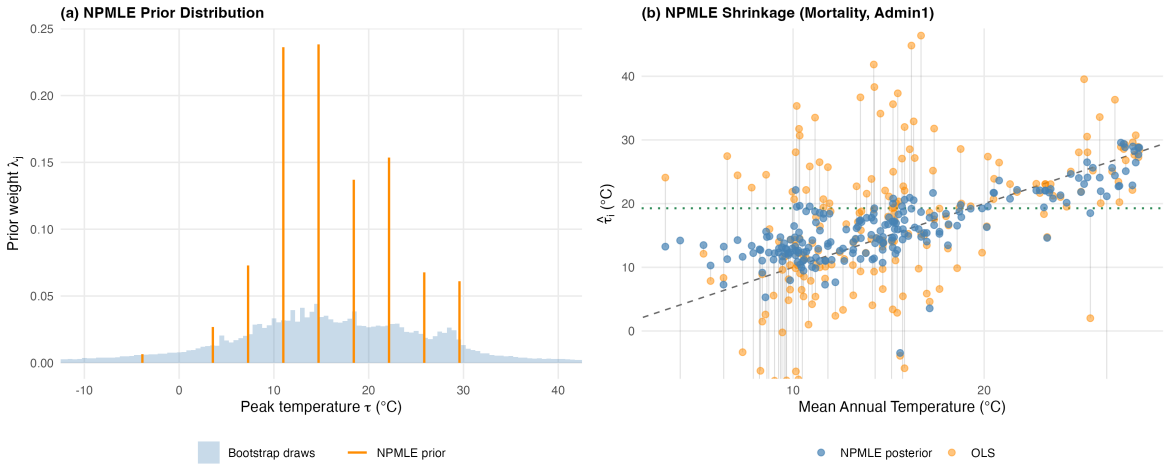


Figure A16: **Estimation of the NPMLE τ_i prior and resulting posterior means for mortality** Panel (a) shows the parametric bootstrap distribution of τ_i draws across all $(\hat{\beta}_{1i}, \hat{\beta}_{2i})$ in blue, and the estimated prior $G()$. The right panel shows posterior means (blue) and OLS point estimates (orange). Little shrinkage occurs.